Iowa State University
Digital Repository

Graduate Theses and Dissertations

Iowa State University Capstones, Theses and Dissertations

2019

Intelligent Team Tutoring: An analysis of communication, cognition, cooperation, and coordination

Kaitlyn M. Ouverson
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Industrial Engineering Commons

Recommended Citation
Ouverson, Kaitlyn M., "Intelligent Team Tutoring: An analysis of communication, cognition, cooperation, and coordination" (2019). *Graduate Theses and Dissertations*. 17535.
https://lib.dr.iastate.edu/etd/17535

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Intelligent Team Tutoring: An analysis of communication, cognition, cooperation, and coordination

by

Kaitlyn Ouverson

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Human Computer Interaction

Program of Study Committee:
Stephen Gilbert, Major Professor
Michael Dorneich
Marcus Credé

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Kaitlyn Ouverson, 2019. All rights reserved.

DEDICATION

As a first-generation graduate student, I did not understand the true depth of the commitment of a thesis until I was deep in writing it. Having no experience with the process, I’ve learned a lot of important lessons beyond just the findings of this study. This thesis represents one of the largest accomplishments of my life – something I can be proud of myself for finishing.

As an efficacious writer, I never expected a 150-page paper would cause me so much stress. Life has a way of introducing new stressors when I am already at max capacity, but I suppose being brought to one’s breaking point is a good way to become more familiar with the things for which you are truly grateful. Thank you to my sister for putting up with me not contributing to the chores, and thank you to my cats, who will never read this, for cuddling me when I needed it most. Thank you to the friends who checked on me when I disappeared for weeks and brought me meals when I said I missed vegetables. And last but not least, thank you to my partner, William, who traveled up from Mississippi to support me when I needed it most. Thank you all for believing in me when I did not believe in myself, and for helping me find the strength to persevere.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my committee chair, Stephen Gilbert, and my committee members, Michael Dorneich and Marcus Credé, for their guidance and support throughout the course of this research.

In addition, I would also like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience. To all of the graduate and undergraduate researchers who worked with me on this project, either developing the environment, writing data-cleaning python scripts, or facilitating data collection with our human participants: thank you Alec Ostrander, Anastacia MacAllister, Anna Slavina, Desmond Bonner, Jamiahus Walton, Adam Kohl, Carter Engen, Colton Goode, Benjamin Halley, Aaron House, Nick Krone, Cole Lingle, Isaac Lewis, Sam Low, Maddie Lynch, Theresa Narog, and Lauren Thilges. I want to also offer my appreciation to those who were willing to participate in the described study, without whom, this thesis would not have been possible.

ABSTRACT

This thesis describes the evaluation of an Intelligent Team Tutoring System (ITTS) designed to teach team and task skills to improve team and individual performance. Previous work has revealed how team communication, shared situational awareness and mental models, and collective efficacy contribute to the success of a team and how these phenomena are molded by the team members' interactions. However, less research has explored the impacts of an ITTS on these dimensions of teamwork. The present study was conducted on 37 teams of three who took on one of two roles – spotter (two people) or sniper – in a military-style task. The teams completed three trials in their original roles, then one spotter and the sniper switched roles in the fourth trial.

Additionally, individuals either received public or private automated feedback from the ITTS on their performance in the task. Results were mixed. Role experience contributed to the mental model or shared situational awareness of that role as it was defined in training, but not to increased similarity of mental models among teammates. Public feedback positively influenced, although only marginally, the percentage of accurately timed communications and was significantly related to lower overall missed communication actions. Individuals' performance was also influenced by the frequency of video game play and the amount of team experience, but only for certain actions. Collective efficacy was impacted by an interaction between experience with cooperative gameplay and frequency of video gaming, where individuals with low gaming frequency but high cooperative gameplay experience had significantly lower collective efficacy than low gamers with no or low co-op experience. Lastly, performance errors were related to individuals' self-reported use of the feedback, in that ignoring the feedback negatively

impacted performance, but selectively following the feedback improved performance. Given previous literature on team dynamics and ITSs, these results are largely unexpected but suggest the feedback style had less impact than was predicted. One team dynamic, collective efficacy, was also shown to be impacted by video game team experience in unanticipated ways, indicating that video game experience and team game experience are indirectly influential to team performance. This research enables the designers of future ITTSs to consider the effects of feedback on coordination and communication tasks more carefully and highlights the importance of the design principle of ensuring a transparent mapping between the feedback and the behavioral triggers that led to it.

CHAPTER 1. GENERAL INTRODUCTION

This work is a culmination of a project originally started in 2014 in collaboration with the Army Research Laboratory's Simulation and Training Technology Center (now the CCDC Soldier Center STTC) as an effort to demonstrate a proof of concept for an intelligent team tutoring system (ITTS). This research was conducted under the approval of IRB #15-399 (Appendix C). Originally created to support team training for two participants in a spotter-surveillance task (Ostrander et al., 2019), the task was expanded in the present effort to include a third teammate with a second role: sniper.

Motivation

Teams and teamwork have been a research interest for nearly a century, beginning with evaluations of work groups in the Hawthorne studies and making way for experiments seeking to understand and improve the use of teamwork within work groups (Bisbey, Reyes, Traylor, & Salas, 2019). Teams, which are traditionally defined as a group of two or more members, each with specific tasks which require coordination of information and activities to reach some common goal or objective (Salas, Dickinson, Converse, & Tannenbaum, 1992), are growing more commonplace, with 80% of work reportedly being team-based by 2010 (Colbry, Hurwitz, & Adair, 2014). Additionally, as the internet becomes more ubiquitous, organizations are making use of broad expertise by individuals at distributed locations through the use of virtual teams, even though research has identified many barriers to their success (Cramton, Orvis, & Wilson, 2007; Haines, 2014; Tong, Yang, & Teo, 2013). For example, these teams may need some assistance in establishing and maintaining professional relationships, especially if they are not familiar with using computer-mediated communication (CMC) (Haines, 2014; Oren & Gilbert, 2010).

In addition to the need for assistance in task skill development, the coordination of tasks and information exchange may also require some measure of training for a team to interact successfully. Team training which encompasses interpersonal skills, or team skills, in addition to task skills is becoming increasingly important as more teams interact primarily virtually; however, training which focuses on interpersonal skills has seen little attention in research (Lane, Core, Gomboc, Karnavat, & Rosenberg, 2007; Orvis, Ruark, Engel, & Ratwani, 2010; Riggio & Lee, 2007). For a distributed team, it is pivotal that the team leverages benefits and minimizes deficits of CMC technology (Alsharo, Gregg, & Ramirez, 2017; Haines, 2014; Pinjani & Palvia, 2013).

Several researchers have posed the question of how best to prepare a team to perform well together. One way is to offer team training which leverages technology to match the environments in which distributed teams will find themselves. Borrowing from the established discipline of Intelligent Tutoring Systems, which see use across various domains (Graesser, Hu, & Sottilare, 2017; Graesser, VanLehn, Rosé, Jordan, & Harter, 2001), Intelligent Team Tutoring Systems (ITTSs) can be used to present information about team and individual performance (Ostrander et al., 2019; Sinatra et al., 2018; Walton et al., 2018) within simulations of dangerous or rare problems or offer virtual training to dispersed teams (Cannon-Bowers & Salas, 1998). Simulation-based training has emerged as an effective way to train teams on the taskwork and teamwork required to reach desired team outcomes (Gorman et al., 2007; Salas, Cooke, & Rosen, 2008).

One advantage of the simulation-based training that ITTSs are built to support is the level of control over the training. In the virtual environment, tasks can be more explicitly controlled, which aids in feedback generation and delivery, an important component of team training (Geister, Konradt, & Hertel, 2006; Salas et al., 2008; Salas, Rosen, Held, & Weissmuller, 2009; Timperley & Hattie, 2007; Walton et al., 2014, 2018). Feedback from the tutor, in turn, can influence teammates' actions, interactions, and shared situational awareness. Indeed, virtual training can be used to train teams on aspects of teamwork, whether that team is collocated and interacting face-to-face or distributed and interacting virtually.

Intelligent Team Tutoring Systems

ITTSs are a relatively new phenomenon based on the success of Intelligent Tutoring Systems (ITSs), which have seen success in computer-assessable subjects such as STEM disciplines (e.g., Craighead, 2008; Graesser, Hu, & Sottilare, 2017; Graesser, VanLehn, Rosé, Jordan, & Harter, 2001; Nye, Graesser, & Hu, 2014). They likely have a future role to play in more complex domains, such as military tasks and healthcare coordination (Graesser et al., 2017). While intelligent tutors have been developed to train team skills, they have infrequently facilitated the training of multiple people simultaneously, instead using artificially intelligent agents as teammates (Buche, Querrec, De Loor, & Chevaillier, 2004; Rickel & Johnson, 1999; Traum, Rickel, Gratch, & Marsella, 2003).

While ITSs have been shown to improve performance across domains, there is not much support for the application of the concepts of these systems to ITTSs. Part of the reason for this gap is the lack of team tutoring systems. This scarcity exists for several reasons, which largely extend beyond the scope of this thesis but are described in detail elsewhere (e.g., Bonner et al., 2016). Finally, there is limited research on the impacts of ITTSs on team performance beyond task execution performance (e.g., Ostrander et al., 2019). While improvements on the task are important to a team's efficaciousness, a good team is not simply one which has gotten the job done. This thesis seeks to address these gaps by looking more closely at team dynamics within an ITTS-facilitated team, and by evaluating the impact of an ITTS on measures of individual and team performance.

The Surveillance with Sniper (SwS) Task and Tutor

This ITTS project was conducted to expand the results of a first proof-of-concept team tutor for a dyad which used the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg, & Holden, 2012) to coordinate feedback to participants who were completing a surveillance task in a serious game engine, Virtual Battlespace 2 (VBS2). The Surveillance with Sniper task (SwS), which is discussed more in-depth in this thesis, was developed to support three participants: two in a spotter role and one in a sniper role. In addition to performance metrics, data were collected to assess the individual and team-level differences in team skills and familiarity with video game and team environments.

This paper explores the relationships between performance in the task, feedback delivery, and the aforementioned individual and team-level differences. The results reveal important considerations for future development of ITTSs, for instance, by illuminating how distinct actions may respond differently to feedback.

Thesis Organization

The following sections will illuminate the background literature important to understanding the problem (Chapter 2), while the specific task and tutor under examination are detailed in Chapter 3. The methods of the study are described in Chapter 4. Chapter 5 features a journal manuscript which analyzes the communication and team cognition variables, while the performance (coordination) and collective efficacy variables are analyzed in Chapter 0. Finally, the thesis concludes with a summary of the unique contributions of this work and where efforts should be focused on moving forward (Chapter 7).

CHAPTER 2. LITERATURE REVIEW

This chapter describes the literature on theories that offer a basis for understanding the remainder of this thesis. First, Intelligent Tutoring Systems (ITSs) are briefly reviewed to offer insight into the origin for Intelligent Team Tutoring Systems (ITTSs). After the history of ITSs, the concept of teamwork ability and how to measure it, including a technique which uses behavioral markers, are introduced. These concepts are important for the generation and delivery of just-in-time feedback, which is utilized in the present study. The section culminates in a review of prominent ITTSs, specifically those which influenced the creation of the ITTS used in the present study. Chapter 3 details that ITTS, the Surveillance with Sniper (SwS) task and tutor.

Intelligent Tutoring Systems

Intelligent Tutoring Systems, or ITSs, have been getting attention for a few decades, with articles published as early as 1986 describing the need to apply artificial intelligence to computer-based tutors and laying out a framework for the burgeoning technology (Tchogovadze, 1986). Since then, ITSs have been developed to teach topics from mathematics and physics to information technology and scientific reasoning, and they have done well in this endeavor. Meta-analyses have shown that ITSs are, on average, just as effective as one-to-one human tutoring (Graesser et al., 2017; Kulik & Fletcher, 2016; vanLehn, 2011), which has already been shown to be more effective than traditional classroom teaching by as much as two standard deviations (Bloom, 1984; Cohen, Kulik, & Kulik, 1982). Two ITSs are particularly worth noting in further detail because of previous research on the relationships between the learner and the tutor: ALEKS and AutoTutor.

One notable ITS designed to teach algebra to high school students is ALEKS (Assessment and LEarning in Knowledge Spaces) (Craig et al., 2013; Falmagne, Cosyn, Doignon, & Thiéry, 2006). This ITS was built on the idea that assessments of mathematical knowledge had been conflated with tests of aptitude; for example, a student can completely understand that $y=x$ is graphed as a 45-degree line which resembles a forward slash, but not understand why that line is correct. Instead, ALEKS uses assessments of *knowledge structures*, or collections of types of problems an individual is able to solve, to guide effective tutoring. Under these knowledge structures are a collection of assumptions about the referenced topic.

For example, Elementary Algebra is decomposed to six types of problems – word problems, coordinate plane plotting, monomial multiplication, factorization of monomials, line graphing, and line equation writing (Falmagne et al., 2006). These six problem types can be mastered in specific orders with some overlap in steps. Domain experts determine these orders, and the complete list is used to determine the different paths to complete domain mastery. These paths can help to specify the current *knowledge state*, or list of problem types mastered and ready to be learned by a student, which is then used to guide further learning. The ALEKS test and tutor is used currently in several hundred learning institutions in the US (Falmagne et al., 2006), including Iowa State University.

While ALEKS works well for teaching mathematics due to its use of knowledge states, it was also appreciated by students who interacted with it (Stillson & Alsup, 2003; Stillson & Nag, 2009). Data in these studies were collected to evaluate effectiveness, but also to evaluate usefulness and helpfulness. Some students were not excited about the ALEKS tutor or were intimidated by the technology, which could have been circumvented with user-centered design and a more-robust tutorial. In general, it was most important that the system positively affected student grades.

Another prominent ITS, AutoTutor, excels not only at teaching computer literacy and Newtonian physics but also at responding in and understanding natural language (Graesser et al., 2001; Graesser, Chipman, Haynes, & Olney, 2005). By using latent semantic analysis, regular expressions and frequency-weighted word overlap, AutoTutor positions itself as a learning partner, pushing the limits of ITSs. While Graesser (2016) admits there are still issues with the tutor, a study which employed the Turing test revealed that the AutoTutor could not reliably be distinguished from a human tutor (Person & Graesser, 2002), which suggests that the ITS can relate to students in the same way as a human tutor. Also, several researchers have begun to focus on the student-tutor relationship explicitly (e.g., Koedinger, Matsuda, Maclellan, & McLaughlin, 2015; Ogan et al., 2012), though they have typically done so using a learning sciences lens rather than a lens based on teamwork research. Thus, even though there is work to be done, AutoTutor is easily identified as a leader in learning innovations.

AutoTutor has inspired many spinoffs, including MetaTutor, which teaches students how to use metacognitive strategies to enhance their learning (Nye et al., 2014). MetaTutor promotes Self-Regulated Learning (SRL), a process in which learners set and aim to attain goals by monitoring, regulating, and controlling their cognitive and meta-cognitive processes (Azevedo, Witherspoon, Chauncey, Burkett, & Fike, 2009). MetaTutor is a biology tutor in content, but more importantly, MetaTutor trains learners to use SRL, with the intention of the skills spilling over into other learning contexts.

While ITSs were traditionally developed to teach subject matter with "correct" answers to students, ITSs are being created to encompass more complex domains, such as communication. INTACT (INtelligent Teamwork ACtion Tool) was created as a tutor for communication skills in airborne command and control specialists, and it was shown to improve participant performance across many measures of communication quality, including decreases in latency and omission from pretest to post-test (Freeman, Diedrich, Haimson, Diller, & Roberts, 2003). Training on communication is an important stepping stone to team training, which is discussed further in the following section.

Team Ability

One of the challenges of an ITTS is assessing team skills, but defining the characteristics of a good team is difficult. There are several skills which are related, but there is not a single method of characterizing all good teams. While one marker of a good team is its ability to accomplish tasks, taskwork is only a portion of the performance equation (Salomon & Globerson, 1989). An effective team accomplishes the goals it was created to achieve through teamwork (Salas, Shuffler, Thayer, Bedwell, & Lazzara, 2015), although by defining team ability with *teamwork*, the core competency is not truly illuminated.

However, teamwork can be broken down further into its composite parts; researchers have identified many over the years (e.g., Driskell, Goodwin, Salas, & O'Shea, 2006; Eduardo Salas et al., 2015). Recently, the literature has been synthesized by Salas and colleagues (2015) into nine dimensions of teamwork which are considered critical to the success of a team. Teamwork success is influenced by the *context* of events and behaviors, its member *composition*, and the *culture* in which it exists. The core processes of teamwork or the elements of which it is composed are *coaching*, *conflict*, *cognition*, *communication*, *cooperation*, and *coordination*. Each of these “critical considerations” are defined and summarized below.

Coaching

Many teams are organized hierarchically, under a leader who provides direction and support to a team (Salas et al., 2015); however, even “flat” teams which do not have a formal hierarchy are affected by the leadership behaviors which are encompassed by coaching. Leadership behaviors, such as the diagnosis and treatment of performance errors, are necessary for team success, and the importance of these behaviors is especially highlighted in virtual teams (Gibbs, Sivunen, & Boyraz, 2017; Hoch & Dulebohn, 2017; Hoch & Kozlowski, 2014; Turel & Zhang, 2010). Some of the leadership behaviors may be conceivably offloaded on an artificially intelligent agent which monitors team performance and offers feedback designed to influence behavior change (Yin, Miller, Ioerger, Yen, & Volz, 2000). Indeed, collaborative software tools may be used to coach group development (Oren & Gilbert, 2010) and teamwork antecedents (Freeman & Zachary, 2018; Ostrander et al., 2019). In the SwS task, the tutor plays the role of coach, which means that it must be able to monitor team performance and offer appropriate feedback. The effectiveness of the tutor in this role is addressed in Ostrander et al. (2019)

Conflict

When picturing a high-functioning team, conflict may not be a part of the image. However, this inevitable part of teamwork has been shown, in some cases, to positively influence performance (Bradley, Klotz, Postlethwaite, & Brown, 2013; Gallenkamp et al., 2012; He, Paul, & Dennis, 2018; Windeler, Robert, & Riemenschneider, 2015). While relational conflict, which arises from differences in values and opinions, may be detrimental, disagreements about the distribution of resources and responsibilities are important to team function (He et al., 2018). Reduction of conflict should occur naturally within a team, and reductions of task conflict leads to increases in shared understanding, or cognition (Salas et al., 2015; Windeler et al., 2015). Because of its tie-in to cognition, conflict was not separately measured in this study. While this core teamwork process certainly occurred within the teams in this study, no hypotheses about conflict were explored.

Cognition

Salas and colleagues (2015) detail the importance of team cognition, which directly impacts a group's ability to work cooperatively toward a common goal. A key factor in team cognition is teammate familiarity, and in fact, this is the first step in the model of group development introduced by Tuckman (1965, cited in Driskell, Salas, & Driskell, 2018). One would expect team members who are more familiar with one another would communicate differently than a team of strangers for reasons including established trust (Peters & Manz, 2007) and understanding of teammate knowledge, skills, and abilities (Salas et al., 2015).

Teams of strangers must talk more to understand one another's needs (Kou & Gui, 2014), which would increase the amount of communication. On the other hand, communication may be negatively impacted if the team operates in a high workload environment, in turn hurting other team behaviors such as back-up behavior (taking the initiative to assist a teammate) (Smith-Jentsch et al., 2015). In general, teammate familiarity has been shown to decrease the amount of communication needed for team success (Espevik, Johnsen, Eid, & Thayer, 2006; Marlow, Lacerenza, & Salas, 2016). Because of this factor, familiarity was assessed in the SwS task.

In addition to familiarity, team members should be aware of the common team and shared situations – who must be at this position and at what time if we are going to succeed? These shared mental models have been shown to be important to virtual and face-to-face team performance (Carpenter et al., 2008; Rentsch & Klimoski, 2001; Windeler et al., 2015). Shared situational awareness (SA) is developed, in part, by consulting a team's shared mental model, which is a more long-term representation of the tasks, subtasks, and roles associated with the team's collective goal (Salas, Rosen, Burke, Nicholson, & Howse, 2007; Stout, Cannon-Bowers, & Salas, 2011; Thomas & Bostrom, 2007). Additionally, common ground (another term for shared SA) represents a team-level dynamic understanding of reality – in this case, of the team's environment and goals (Kruijff, Janíček, & Zender, 2012; Sætrevik & Eid, 2014; Salas et al., 2007). Shared SA can be developed through role cross-training as well as communication (Gorman et al., 2007).

Communication

Communication is recognized as important to teamwork success. It influences teammate rapport (Nardi, 2005) and belongingness (Haines, 2014), which have both been tied to team performance. In a virtual team, in which members interact primarily through the use of computer-mediated communication (CMC) technologies, communication emerges as one of the most important pieces of teamwork (Charlier, Stewart, Greco, & Reeves, 2016; Tong et al., 2013).

Through communication, teams develop *common ground*, or a team-wide understanding of the team's task and goals, (Driskell et al., 2006; Oren & Gilbert, 2011; Paul & Ray, 2014) and reduce *situation invisibility*. Situation invisibility occurs when information regarding an individual's context is not adequately communicated, which results in increased dispositional, rather than situational attributions of teammate shortcomings, e.g., "That teammate is lazy" vs. "That teammate has slow Wi-Fi" (Cramton et al., 2007). This phenomenon is most prevalent in distributed teams, whose communication does not include context cues unless they are explicitly stated (Cramton et al., 2007).

Communication is closely linked to team cognition, and in a virtual environment, a software-based tutor could help both cultivate proper communication and fill in gaps when deficits occur. This core teamwork process is measured in the present study to examine the influence of the tutor on team communication. Acknowledging previous literature (Espevik et al., 2006; Marlow et al., 2016), the link between prior teammate familiarity and communication is also studied in this work.

Cooperation

An additional factor which could reasonably be expected to impact communication within a team is the collective efficacy of its members. For Salas and colleagues (2015), collective efficacy falls under the critical consideration of *cooperation*, which involves the attitudes, beliefs, and feelings that motivate individuals to work effectively in teams. Collective efficacy is an individual’s belief in abilities of the self and others in team settings to work together toward accomplishing some goal (e.g., Bandura, 1982; Chou, Lin, Chang, & Chuang, 2013; Porter, Gogus, & Yu, 2011; Salas et al., 2015; Stajkovic, Lee, & Nyberg, 2009). When individuals believe their efforts will be fruitful, they are more likely to exert additional effort (Bandura, 1986; Chou et al., 2013). For example, a student finding art enjoyable and easy to do may be more willing to practice sketching still-life figures.

In the same way, if a person believes the team is bound to succeed, that team member will spend more time to fulfill the duties of the role, including communicative duties. Further, research has shown a link between collective efficacy and performance feedback (Porter et al., 2011; Tasa, Taggar, & Seijts, 2007), which is to say that as individuals recognize that their team has been performing well, their opinions of the team’s ability improves, which increases their teamwork and leads to continued success. Other research has additionally shown self-efficacy about a task is directly related to performance in that task

(Locke & Latham, 2015). Because of this, collective efficacy is examined within this thesis; specifically, the ties between collective efficacy and current performance as well as past experience are studied.

Coordination

While team ability is not solely defined by task completion, this aspect of teamwork is important to *coordination*. Coordination is defined by Salas and colleagues (2015) as the pulling together of each member's resources to complete the team's tasks. Coordination is heavily reliant on communication and team cognition. If a team's members are not communicating their needs, coordination is still possible through the activation of shared mental models and shared SA, i.e., by utilizing back up behaviors (Bradshaw et al., 2008; Chang, Hung, & Hsieh, 2014; Freeman & Wohn, 2018; Paul, Drake, & Liang, 2016; Porter et al., 2011). For a task to be considered a team task, the individual parts need to be interdependent, so that proper communication and coordination is necessary for optimal performance (Bonner et al., 2017; Gilbert et al., 2017). Within the present study as well, success in the developed tasks is reliant upon proper interdependent action.

The Critical Considerations and the SwS

Each of these critical considerations of teamwork are present, to some degree, in all teams. As such, they each appear in this study; however, only four (communication, cognition, cooperation, and coordination) are examined within this work. Task conflict is relevant to this sort of short-duration team (He et al., 2018), but as the task is guided by the tutor, conflict was expected to be minimized. The phenomenon of disagreements in task and goal prioritization can further be assessed using measures designed to capture the team members' cognition, so any conflict examinations would be redundant. Coaching would be relevant to this study, since an artificially intelligent agent serves as a guide to the teams;

however, questions related to the coaching ability of the tutor are better addressed in a tutor development paper (i.e., Gilbert et al., 2017; Ostrander et al., 2019) and are outside the scope of the present work. The four critical considerations are summarized in Table 1.

Table 1. Definitions of the critical considerations of teamwork under examination in this paper, as popularized by Salas et al. (2015).

Team Dimension	Description
Communication	The information exchange among team members which guides teams to a common goal and a common understanding of that goal.
Cognition	The familiarity of the members in a team and their shared understanding of the team's tasks and roles in completing them.
Cooperation	The beliefs of the members of the team which motivate teamwork behavior.
Coordination	The act of transforming team-member resources into team outcomes.

Measuring Team Ability

The analysis of teamwork can be accomplished by the use of event-based measurement, either based on behavioral markers (e.g., Sottilare et al., 2018) during the task, by self-report after the completion of the task, or by analyzing program-collected or quiz-collected data after the experiment. As in most cases, self-report measures are inherently biased, and for team ability, the same is true. Ostrander et al (2019) demonstrated that actual performance correlates with self-reported performance infrequently. A more salient issue with self-report measures is the inability to use it in automatic assessment of teamwork. For this, event-based measurement is more appropriate (Fowlkes, Dwyer, Oser, & Salas, 1998). Indeed, event-based measurement was utilized in directing the ITTS's feedback in the present study and in measuring participant performance.

Event-based measurement was used to identify teamwork abilities in vivo. By defining quantifiable markers for these behaviors before evaluating performance, feedback was generated and given by the ITTS. These behavioral markers were derived from the constructs of team performance that are important to attaining positive team outcomes, such as those identified by Salas and colleagues (2015) and explained in the previous section. These behavioral markers helped to create a model of team performance. For example, individual task error rates were used in the present study to evaluate and diagnose coordination.

Additionally, communication may be measured by volume, frequency, timing, and content of verbal utterances by individuals, either using observation or tools such as the sociometric badge (Pentland, 2012; Wiese, Shuffler, & Salas, 2015). Communication events, which may include button presses which have pre-established meaning, may also be useful as a behavioral marker. Within the present study, communication events rather than verbal utterances were measured for this purpose.

However, not all aspects of teamwork can be measured using event-based metrics. For teamwork components which are mental in nature, such as cognition or cooperation, self-report can be useful despite its shortcomings. In the present study, cognition was measured via quizzes that assessed understanding of the task and roles associated with the SwS, modeled after the assessments of shared mental models and situational awareness used by Sætrevik & Eid (2014). Another teamwork component examined in the present work, cooperation, was operationalized and measured as collective efficacy (Salas et al., 2015). While these could not be measured and directly improved using feedback supplied by the tutor, the effects of a tutor on such components of teamwork were evaluated after a tutoring session. Indeed, while the SwS did not use cognition or cooperation to evaluate users, these teamwork components were used to examine the impacts of the tutor on team dynamics.

Feedback

Feedback is recognized as an important consideration for the development of an ITS and an ITTS (Walton et al., 2018). It has been established that proper feedback, which is aimed at increasing awareness of the task or of the process, or at increasing self-regulation, has positive impacts on learning (Azevedo, Witherspoon, Chauncey, Burkett, & Fike, 2009; Gilbert et al., 2017; Timperley & Hattie, 2007). Feedback's effectiveness is further mediated by goalsetting, as shown by Locke & Latham (2015). If feedback does not direct a person toward their goals, it does not have as big an effect on that person's performance.

Beyond just offering team members a way to improve performance, feedback affects the way team members receive an ITTS. Inappropriate etiquette (Dorneich, Ververs, Mathan, Whitlow, & Hayes, 2012; Walton et al., 2014; Yang & Dorneich, 2018) or excessive messages (Price, Mudrick, Taub, & Azevedo, 2018) have been shown to have a negative impact on learner performance. In the case of excessive messages, negative affect reduces the learner's willingness to learn, in turn impacting their performance. In a team setting, appropriate etiquette becomes more important, as the tutor must maneuver human-human interaction in addition to game-state considerations, such as taskload related to the number of visual-search targets on-screen.

Team training presents additional unique feedback considerations. Gilbert and colleagues (2017) discuss feedback design and its impact on learning, noting several considerations which go into making and giving feedback. There are pedagogical considerations: What skills are necessitated by the team's roles and should be used to guide feedback? What is the best way to give teamwork feedback? The answers to these questions depend largely on the context of the team. In a virtual environment, tasks can be more explicitly controlled, which aids in feedback generation and delivery, important components of team training (Geister et al., 2006; Salas et al., 2008, 2009; Timperley & Hattie, 2007; Walton et al., 2014, 2018). A tightly controlled team task has specific outcomes and may even include specific team challenges, as in the SwS, therefore allowing for the definition of events for which feedback can be supplied. Beyond the content of feedback – its pedagogy and affect – it is important to consider how to address and deliver feedback, and what behaviors to assess when crafting said feedback.

In many ways, this is a matter of individual versus team-level feedback. These levels have at least three dimensions, which are highlighted in Figure 1. Behavior assessment at the team level means that feedback is generated for team behaviors, like coordination and communication, while individual assessment measures individual tasks, like the speed of communication and accuracy of responses. Feedback delivery refers to the difference in the recipient of the feedback, either the individual for whom it is relevant or the whole team. The last consideration for feedback design is the audience to whom the feedback is addressed,

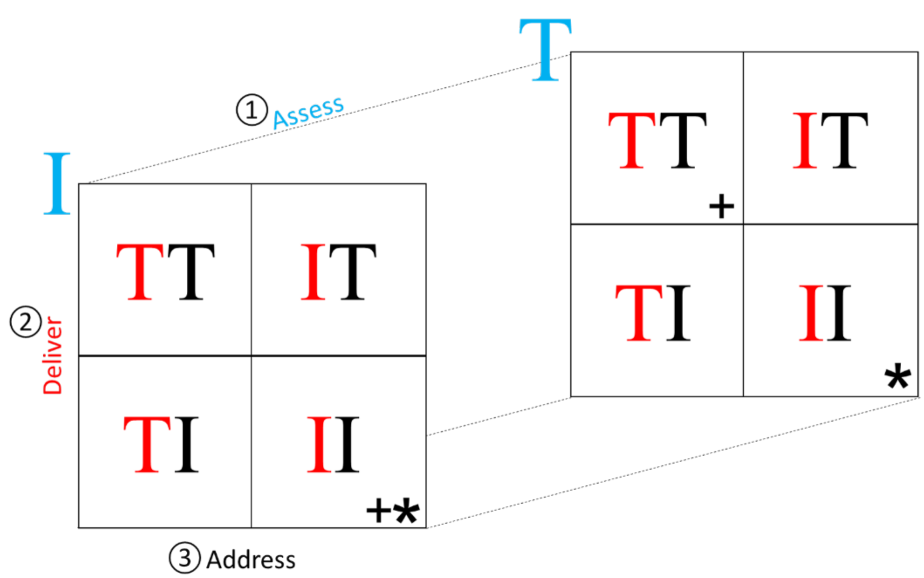

*Figure 1. The three dimensions of feedback considerations: **1** Assess, **2** Deliver, **3** Address. * identifies comparisons made in this research, + identifies comparisons made in previous research by this team (Ostrander et al., 2019).*

either to the team or to each individual. Not all combinations of these considerations make sense, for example, feedback which is gathered from an assessment of the Team, delivered to the Individual, but addressed to the Team. However, the performance and team outcomes related to few combinations have been explicitly explored.

Team Training

Team training, as conducted by ITTSs, can be conceptualized in at least two ways: either as the work of several automated individual tutors which communicate with one another and with their individual human trainees (as described by Sottilare, Holden, Brawner, & Goldberg, 2011) or as a job for a single, omniscient team tutor. An all-knowing team tutor would coordinate information about each user and the team to give feedback. ITTSs and ITTS-like systems have been created which follow both molds, while additionally featuring different choices on feedback timing (after-action or just-in-time), team composition, agent role, and the feedback dimensions of assessment, address, and delivery highlighted in Figure 1.

One of the first ITTS-like systems was the Advanced Embedded Training System (AETS), which facilitated Naval air defense training (Zachary et al., 1999). The AETS monitored the learners' button presses, speech, and eye movements to supplement the work of a human trainer. The human trainer's time focused on aggregating data from the AETS into a team-level after-action review, while automated task feedback was given just-in-time by the AETS to individual learners.

In the AETS, team members were assigned specific jobs, and feedback on performance was given by both the software agent and the human trainer. In the Team Multiple Errands Task (TMET; Walton et al., 2015) the software agent, or tutor, supplied real-time individual *and* team-level feedback to a team of three as they completed a multiplayer virtual shopping task. The TMET extends a classic single-person shopping-based cognitive task to a team of three.

The team member roles required by a team task often play an important role in the dynamics of the team. In TMET, the team members' roles were homogeneous, with no specific job roles or background training. In education and the workforce, team members often play different roles. Software development teams, for example, consist of designers, engineers, and social scientists who work together to ship new products. While homogeneity makes the study of a team simpler and more controllable, the ecological validity of the team tutor in such situations is decreased. The SwS team tutor oversees a heterogeneous team of three with members each taking on one of two roles.

While the TMET tutor was not embodied or personified, and all feedback was given as brief phrases or data visualizations based on individual performance or team scores, a different kind of team tutor, Avis (Kumar, Ai, Beuth, & Rosé, 2010), was able to give feedback to a team through conversational dialogue, acting as a guide for learning underlying concepts of mechanical engineering. While Avis could be considered an ITTS, the tutor did not provide feedback for the team as a whole; it instead attended to each learner’s conceptual understanding. Without team-level feedback, team skills are not actively trained, and therefore, Avis could be referred to as a socially capable tutor. Further, the use of teams and conversation by Avis was used to facilitate the learning of the material, rather than the improvement of team skills.

The agents developed within the Programme for International Student Assessment (PISA) 2015 (Fiore et al., 2017) engaged learners in conversation in much the same way as Avis. Instead of tutoring as a facilitator, the agent (and sometimes multiple agents with various skills) worked collaboratively with the learner, as a peer to solve problems. In this way, the PISA 2015 agents are not tutors, although their goals are similar to agents in tutoring roles. Additionally, tasks always involved only one human teammate, rather than a multiple-human team. While the PISA 2015 encouraged the use of soft skills, which are skills such as the reading of body-language and other back-channels of communication, in engaging conversations, the inclusion of only one human per team limited the applicability of their results to teams of more than one person. The SwS team tutor, while not socially capable or conversational like Avis and the PISA 2015 agents, focuses on fostering improvement in the soft skills of the team, specifically timely and accurate communication and appropriate acknowledgments of received communication.

Unlike Avis and the PISA system, the SwS (as well as AETS and TMET) is a fast-paced, high-cognitive load psychomotor performance task that requires steady focused attention and does not typically allow for conversational dialogue. Although SwS is not conversational, it is important to consider some of the ways in which the tutor functions as an intelligent agent within a human-agent team. For instance, the tutor, which facilitates training by giving advice on how best to operate as a team, can be identified as working from a supervisory role, as opposed to the helper role taken by the agents behind the Avis and PISA 2015 tutors (Ouverson, Pena, Walton, Gilbert, & Dorneich, 2018). It is important to recognize the users' mental models of the tutor: it cannot be simply considered a software program. It is instead an agent in conversation with the users. To facilitate change in users of an ITTS or an ITS, the agent must be persuasive, maintaining the attention of its learners using immediacy cues, socio-emotional behaviors, and etiquette, when possible (Chidambaram, Chiang, & Mutlu, 2012; Dorneich et al., 2012; Lohani, Stokes, Mccoy, Bailey, & Rivers, 2016; Yang & Dorneich, 2016). Additionally, the agent must motivate users to learn and improve by appropriately utilizing strategies such as social comparison feedback and praise (Mumm & Mutlu, 2011).

Before delving completely into the SwS team tutor, it is worth noting its predecessor, the Surveillance Team Tutor (STT). The military task for STT was developed in Virtual Battlespace 2, a serious game engine. In the task, two spotters were positioned on top of a building in the middle of a virtual village environment which included walls as obstacles between which OPFOR (OPposing FORces) could run. Each spotter was responsible for watching a zone consisting of half of the environment and alerting his or her teammate to OPFOR who neared that teammate's zone. The full task consisted of a *Transfer* event, in which one spotter alerted the other to an approaching OPFOR; an *Acknowledge* event, in which the receiving spotter acknowledged the transfer; and an *Identify* event, in which the second spotter noted receiving the OPFOR into his or her zone. Teammates passed this information to each other via a verbal communication channel and to the tutor using pre-assigned keyboard keys.

The STT offered two types of feedback to its users, which were referred to as Team feedback and Individual feedback. Team feedback followed a TTT configuration, where assessment, delivery, and address all occur at a team level, while Individual feedback followed an III configuration (Ostrander et al., 2019). These two were chosen because, as shown in Figure 1 and detailed in Chapter 5, they represent the two most different configurations for feedback in this three-dimensional space of feedback characteristics. In the STT, participants in a control condition were given no feedback.

While the tutor was shown to have limited impact on the performance of participants and their teams (Ostrander et al., 2019), there were promising results related to the impact of feedback on performance, shared mental models, and overestimation of performance. Feedback (as compared to no feedback) on the Acknowledge subtask did result in fewer errors, and Team feedback (as compared to Individual feedback and no feedback) reduced the tendency of individuals to rate their teammates as having performed poorly. Lastly, in the Individual and No feedback conditions, participants' self-ratings of individual performance did not correlate significantly with their tutor-assessed performance.

It is worth noting that the task was created to mimic a video game environment. Cooperative video games, or games that players team up to play, are common, as evidenced by Steam's Top games by current player count, which shows that of the top ten games, nine are tagged as "multiplayer" (Valve Corporation, 2019). While these multiplayer games include a mix of cooperative and competitive gameplay, there are none that do not allow players to team up and play cooperatively. Other game companies have also boasted a large interest in cooperative video games, hosting and promoting e-sports, such as Blizzard Entertainment's Overwatch League and Riot Games' League of Legends Championship Series. Another of Blizzard Entertainment's games, a massively multiplayer online game called World Of Warcraft, has been used in previous work to understand team dynamics (Benefield, Shen, & Leavitt, 2016; Billieux et al., 2013; Dabbish, Kraut, Patton, Heinz, & College, 2012). The players in these games must learn to be good teammates if they want to succeed, and by being good teammates, the game itself becomes more appealing (Billieux et al., 2013). In the SwS study, participants were asked about their experience with team-based video games, as this experience might have affected their performance and impressions of the team.

The present study examines an ITTS which was built upon the foundation laid out by the STT. The Surveillance with Sniper (SwS) tutor was created using GIFT and operated in an environment also built in VBS2. The original role is referred to as the spotter role and still required these teammates to communicate with one another as they did in the STT. In SwS, spotters additionally introduced the task’s OPFOR positions to the third team member who used a sniper scope to assess the level of threat posed by each of the targets. The full sequence of this task and the details of the tutor are described in the next chapter.

CHAPTER 3. THE SWS TASK AND TUTOR

This thesis describes a study examining the Surveillance with Sniper (SwS) team tutor, which is a continuation of previous work described by Ostrander and colleagues (2019). In that previous work, the Surveillance Team Tutor (STT) was created in collaboration with the Army Research Lab as the first ITTS which used just-in-time feedback to guide dyads through a military-style task. Feedback was supported via the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare et al., 2012), which was modified from its original form as a framework for supporting single-user ITSs to facilitate tutoring for pairs of users. This task and the accompanying tutor are described in detail below.

Surveillance with Sniper: Task and Tutor

The Surveillance with Sniper (SwS) Team Tutor was created in collaboration with the Army Research Laboratory in much the same way as STT. However, there are marked differences between the two tasks and the two tutors. Namely, SwS incorporates a second role and third teammate, which the tutor was expanded to accommodate.

Team Task

In the SwS, there are two spotters, still atop a building in a village, each responsible for surveillance of their respective zones. In addition, there is a Sniper, who is positioned in a side tower that can see the entire village. The sniper's duty is to assess whether a running person identified by the spotters is a threat.

The task dynamics are described in the following example. Spotter 1 notes two targets approaching on the side of the middle building with one pole by striking the "1" key twice

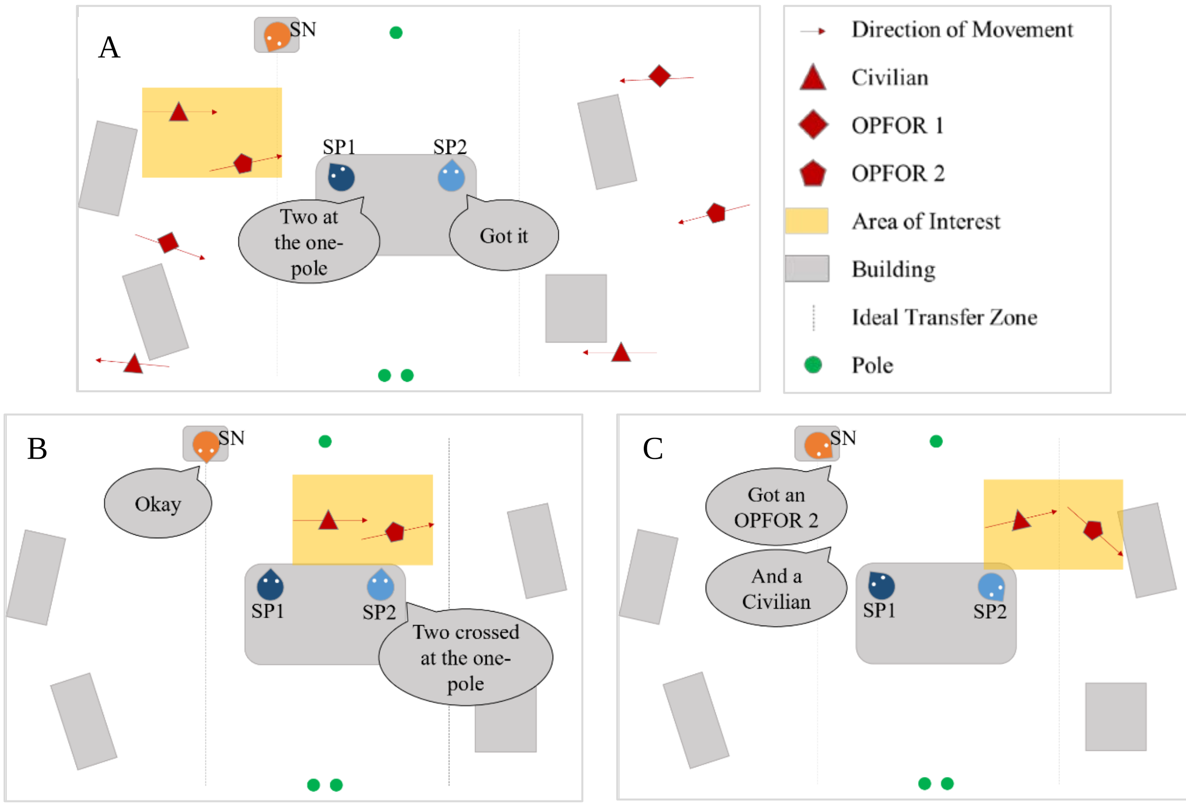

Figure 2. *The SwS task within the simplified environment.* ***A)*** *Spotter 1 (SP1) Transfers two targets to Spotter 2 (SP2) as they approach the center building. Spotter 2 Acknowledges.* ***B)*** *Spotter 2 Identifies the targets as they cross on the side of the single pole and alerts the Sniper (SN).* ***C)*** *The Sniper Assesses the targets as a civilian and a level-2 OPFOR.*

and *Transfers* them to Spotter 2, saying “two at the one pole!” (The other side of the environment is denoted by two poles and would have been identified using the “2” key and a corresponding utterance.) Spotter 2 *Acknowledges* that communication from Spotter 1 both by speaking aloud and double-pressing the “E” key. Spotter 2 then *Identifies* the OPFOR for the Sniper by saying words like, “two crossed at the one pole, Sniper,” alerting him or her to their presence, while pressing the Spacebar twice, once for each entity. The Sniper then *Acknowledges* that communication with two strikes of the “E” key and aloud. To finish the sequence, the Sniper *Assesses* the threat posed by each of the targets, either designating them a civilian (“Z” key), a level 1 OPFOR (“X” key), or a level 2 OPFOR (“C” key). The keystrokes (*1*, *2*, *E*, *SPACE*, *Z*, *X*, and *C*) were used to record each action with the tutor.

These keystrokes were chosen so that a player could type them all with the left hand while using the right hand to control a mouse to pan through the environment. This sequence is shown in a simplified version of the task environment in Figure 2. The task environment and GIFT tutor interface seen by participants in the present study is shown in Figure 3.

Tutor Architecture

The SwS team tutor architecture was an expansion of the two-person team architecture from STT. An exhaustive description of this base architecture was given by Gilbert and colleagues (2015). Because GIFT was created as a platform for one-to-one individual tutoring, the architecture to support team tutoring was a complicated integration of individual monitoring and compilation of team data.

First, each participant's performance was tracked by GIFT, running as a local server on each computer. The GIFT Gateway Module communicated with VBS2 to record both team and individual data in the form of short-term information updates, such as learner

Figure 3. Image of the SwS environment as experienced by participants. Image originally included in the spotlight video filmed by Iowa State University College of Engineering (2018)

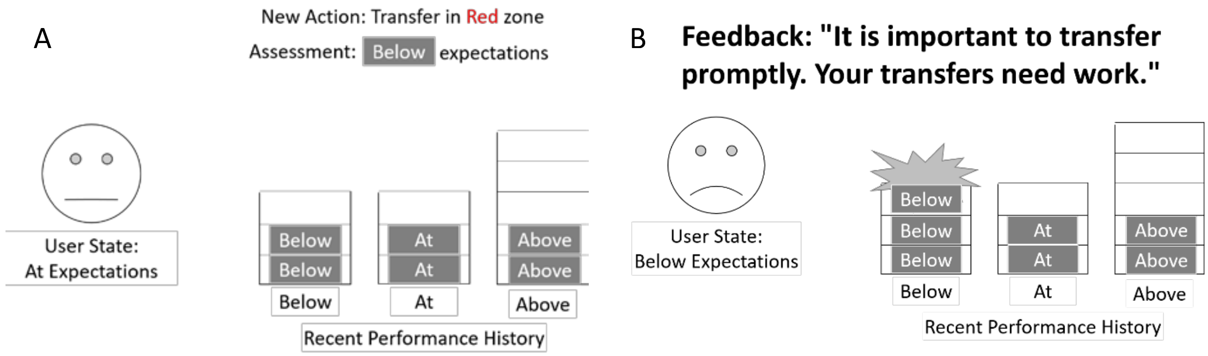

Figure 4. Feedback was not given after ever player event, but rather after "buckets" of similar player events were filled. After a new user state was identified (B), players were given constructive feedback on their recent performance. Adapted from figure originally featured in Ostrander et al. (2019).

actions and environment events, and long-term stored information about those learner actions, for example, including information about how users improved over time and what feedback and feedback triggers were present.

In the long-term storage modules, the GIFT Domain Module contained the feedback and the conditions by which that feedback was to be triggered or not triggered, while the GIFT Learner Module kept track of individual and team performance. For SwS, the Domain Module contained information about when to give feedback for the Transfer, Acknowledge, Identify, and Assessment subtasks, while the type of feedback (either Private or Public) was determined by the tutoring paradigm which was selected upon task set up. Subtasks were evaluated as either *Below Expectation*, *At Expectation*, or *Above Expectation* according to specifications shown in Figure 4. To avoid having feedback triggered too frequently (by every player action), a "bucket" system was used. When the requisite number of actions were documented at a single level of performance (i.e., when the Below, At, or Above stack is filled), feedback was triggered, and the user state moved accordingly, if appropriate. This feedback trigger adjustment was further detailed by Gilbert and colleagues (2017).

Feedback given to the Spotters and Snipers is exemplified in Table 2. The Public feedback condition, in which each team member's actions trigger feedback given to the whole team, would have normally resulted in participants receiving more than twice the feedback as those in the Private condition, in which feedback only goes to the team member whose action triggered it. To balance the amount of feedback received in the Public condition, feedback was not given for Transfer events. Since the Acknowledge actions are directly tied to the Transfer or Identify actions, this feedback is framed as being triggered by the Transfer-Acknowledge or Identify-Acknowledge pairs in the Public condition.

Table 2. Examples of feedback given for each player action in the task, after the relevant "bucket" is filled, and the condition in which the feedback is present.

Task	**Feedback**	**Condition**
Transfer	*"It is important to effectively communicate crossings"*	Private
Acknowledge	*"It is important to confirm at appropriate times"*	Private
Identify	*"It is important to identify targets as quickly as possible"*	Private and Public
Assessment	*"Remember to assess the threats posed by all crossing targets"*	Private and Public
Transfer-Acknowledge Pairs	*"Acknowledge your communications as soon as you receive them"*	Public
Identify-Acknowledge Pairs	*"It is important to confirm at appropriate times. Your team communication needs work"*	Public

All feedback was given in response to the comparisons between the player actions and the relevant GIFT domain knowledge file (DKF), wherein individual tasks were evaluated against the three individual-level DKFs in the Private feedback condition and team tasks were evaluated against the one team-level DKF in the Public feedback condition. In some sense, this can be interpreted as the having three individual tutors in the Private condition (one per teammate), and having one overall team tutor in the Public condition.

Further, although feedback was given in response to task performance, the tasks were designed to require interdependence so that feedback would directly point participants to better teamwork. Feedback was focused on developing the team skills of coordination and cognition by identifying performance on task-based behavioral markers of those skills.

Summary

This thesis describes research which expands upon the tradition of intelligent tutoring systems by contributing knowledge about the impact of ITTSs on team and individual performance, communication, cognition, composition, and coordination. Team training by ITTSs is one solution to the training needs of an increasingly virtual world of work, made up of distributed teams. Work has so far been done to understand the important components of teams and teamwork, and the influence each of these have on team performance. Limited studies have been conducted which look at the relationships between agent-guided team training, as exists in an ITTS, and teamwork. This thesis represents a step toward a more complete understanding of the impact of ITTSs on effective team performance and teamwork. The following chapter describe the specific methods used to evaluate the SwS team tutor on these dimensions.

CHAPTER 4. METHODS

Experiment Overview

The experiment was conducted over four five-minute trials, as detailed in Table 3. The feedback privacy was manipulated between-subjects, and therefore between-teams. Additionally, in the final trial (Trial 4), one of the spotters (randomized for each team) switched roles with the sniper, meaning that the performance of the spotters on measures of cognition could be compared to assess cross-training effects. Additionally, this meant that there are three levels of role naïveté: *Total naïveté* (Trial 1), *Experienced* (Trials 2 and 3), and *Partial naïveté* (Trial 4, Secondary Sniper and Spotter). The partial naïveté references that the person playing sniper in Trial 4 had never performed that role before but should be generally aware of what tasks might be included based on experience as a spotter interacting with the sniper. The same partial naïveté refers to the experience of the new Trial 4 spotter, as well. For clarification of roles below, team members who served as a spotter in Trials 1-3 and then switched to sniper are called "primary spotters," and team members who served as a sniper in Trials 1-3 and then switched to spotter as called "primary snipers."

Table 3. The experiment was structured such that each team experienced the task and tutor four times, with each team receiving either private or public feedback.

Feedback Condition	# Teams	Trial 1	Trial 2	Trial 3	Trial 4 (Role Swap)
Private	19	Task (5 min)	Task (5 min)	Task (5 min)	Task (5 min)
Public	18	Task (5 min)	Task (5 min)	Task (5 min)	Task (5 min)

This chapter begins with a presentation of the research predictions, and then enumerates the response and explanatory variables. Finally, a description of the data analysis procedures introduces the transition into the journal paper described in Chapter 5 and the additional findings in Chapter 6.

Research Predictions

There are several hypotheses which emerge after reviewing the literature. Hypotheses 1 through 6 are examined in Chapter 5, while the Hypotheses 7A through 15 are examined in Chapter 0. The hypothesis are presented below in a non-numerical order which makes the most sense for overall examination, but they are analyzed according to the response variables of interest in Chapters 5 and 6.

Shared Situational Awareness

Teamwork ability is often quantified using conversation-based team metrics; however, these are not as applicable to the analysis of teamwork within the SwS task. Researchers have identified measures of teamwork beyond conversation-based metrics (Salas et al., 2009; Wiese et al., 2015) that can help identify teamwork abilities in vivo, where outcomes are not directly quantifiable or occur over an extended time period. Salas, Shuffler, Thayer, Bedwell, and Lazzara (2015) synthesized research on team success metrics in an attempt to standardize the terminology and direct future research. Of particular importance to the current research effort is their concept of Cognition. Salas and colleagues (2015) detail the importance of team Cognition, which describes a group's ability to function cooperatively toward a common goal. To do so, all team members should be aware of the common team and shared situations – who must be at this position at what time if we are going to succeed? In this way, team Cognition is merely a superset of shared situational awareness.

This work investigates how the amount of role experience influences shared situational awareness (SA). Sætrevik and Eid (2014) note the importance of SA for each member of a team, stating that for a team to perform its best, each member must understand their tasks within the team. This understanding is developed through training and experience with the task. Therefore, it is hypothesized that:

> H1: For the participants who experience more than one role, shared situational awareness will be higher, compared to participants who experience only one role.

Feedback

In the first iteration of the original STT, feedback derived from team assessment, delivered and addressed to the team (TTT configuration) was compared to that derived from individual assessment and delivered and addressed to the individual (III configuration). The present research seeks to uncover the difference between ITT and III configurations by examining the differences in shared SA and communication. The researchers expect that feedback which is given to the whole team would have a positive effect on the shared SA since the use of a team-level delivery feedback dimension will result in more shared information than individualized feedback. Additionally, whole-team feedback would allow teammates to keep track of their whole-team communication performance, correcting as necessary. It was hypothesized that:

> H5: Public feedback will result in higher shared situational awareness than private feedback.
>
> H6: Public feedback will result in lower communication errors than private feedback.

It was expected that by making feedback public, team-level performance, in addition to communication performance and shared SA, would be improved (Alsharo et al., 2017; Cramton et al., 2007; Sætrevik & Eid, 2014; Windeler et al., 2015). Research has also shown that privately directed feedback works best for improving individual performance (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Mumm & Mutlu, 2011), and the same was expected here. Lastly, individual performance was also expected to be improved by participant use and appreciation of the feedback, since feedback will not affect the performance of a person who never looks at the feedback. It was hypothesized that:

> H7A: Teams receiving public feedback will have higher team performance than teams receiving private feedback.
>
> H7B: Individuals receiving private feedback will have higher individual performance than those receiving public feedback.
>
> H8: Participants who use or find the feedback more helpful will have higher individual performance than those who do not use the feedback or do not find it as helpful.

Previous Experience (with teamwork and cooperative-play video games)

Previous experience in teams across domains should logically impact an individual's future performance within a team. Research points toward the impacts of previous teamwork experience on future teamworking attitudes (Rudawska, 2017) and the understanding of how to perform well on future teams (Hirsch & Mckenna, 2008; Reagans, Argote, & Brooks, 2005; Rentsch, Heffner, & Duffy, 1994). The tie between experience and performance has, thus far, been indirect; an individual's experience has influenced his or her schema of teamwork, which in turn influences that person's performance in a team setting. Because of this, the researchers hypothesize that:

> H3: Persons who work in teams more often will have higher shared situational awareness than those who work in teams infrequently.

Given that collective efficacy is established over time based on the experiences in team settings, it was expected that previous experience playing cooperative, multiplayer video games would impact the collective efficacy of the participants. It was also expected that people who frequently play video games would have higher performance in the task since the setting of the experiment was a video game environment. Therefore, the researchers hypothesized that:

> H9: Persons who play a higher proportion of co-op video games and play video games more often will have higher overall collective efficacy than those who play infrequently and in fewer co-op games.
>
> H10: Participants with more frequent video game experience will have higher individual performance than those with less experience.

Similarly, general experience working in teams should influence collective efficacy (Tasa et al., 2007). More frequent experience with teams also means more practice orienting oneself to the role requirements of team members. This experience in turn facilitates positive team outcomes, such as higher performance. Because of this, the researchers hypothesized that:

> H11: Participants with more frequent team experience will have a higher overall collective efficacy than those with less frequent experience.
>
> H12: Participants with more frequent team experience will have higher individual performance than those with less experience.

Familiarity (with teammates, role and task)

Familiarity with teammates can be expected to influence performance. Familiar teams do not need to communicate as often about trivial matters and can anticipate the needs of their teammates (Smith-Jentsch et al., 2015). In this way, performance should be higher for teams and teammates who are more familiar with their team members.

Another way to understand the abilities of a team may be to examine the familiarity of its teammates with one another. Friendship has been shown to improve performance (Mason & Clauset, 2013), but familiarity does not necessarily need to mean a relationship on the level of “friends.” Familiarity on a professional level fosters an understanding of teammates’ skillsets and strengths, thereby increasing the effectiveness of a team, especially under high workloads (Smith-Jentsch et al., 2015). Lastly, teammate familiarity offers an easy path to belongingness, which is an important component of team performance (Haines, 2014). An individual who knows at least one of his or her teammates should perform better on the team, communicating more with teammates. It is therefore hypothesized that:

> H2: Participants who are familiar with at least one teammate will have fewer Acknowledgment errors than those who are not familiar with any teammates.
>
> H14A: Participants who are more familiar with their teammates will have higher individual performance than those who are less familiar.
>
> H14B: Teams with members who are fully familiar with one-another will perform better than teams with no and partial familiarity.

Similar to experience with teamwork, experience with one’s role is important to team performance. Reagans and colleagues (2005) found that experience within a team, and with one’s role on that team, impacted team performance. Experience with a team also has been shown to influence the team’s shared mental models which have been shown to decrease the need for communication (Carpenter et al., 2008). However, the present study used feedback to actively encourage team communication, so it is expected that communication would be unaffected by the establishment of shared mental models.

Familiarity with role is also expected to influence performance. Team members who are not as familiar with their role will perform more poorly. This decrease in performance is expected because individuals must adapt to their roles. As participants gain experience in the task, under the guide of a tutor designed to improve performance, their collective efficacy should improve. However, even if participants have familiarity with the overall task, they may not have the experience with the specific role to know the best ways to complete their role’s subtasks effectively. Since a role swap occurs in the last trial of the study, participants will be less familiar with their roles in Trial 1 and in Trial 4 after a role swap than they are in the middle two trials.

In the fourth trial of the present study, two participants on each team experience a new role. Role switching establishes a grounded understanding of the team's roles and has been shown to foster more effective communication (Sottilare et al., 2011). However, the immediate effect of the role switch is expected to increase task load and decrease communication. Therefore, it is hypothesized that:

> H4: In Trials 1 and 4, Acknowledge errors for participants will be higher than in Trials 2 and 3.
>
> H13: Spotters who do not switch roles and therefore are fully experienced in that role will have higher individual performance than those who switch into a role and experience partial naïveté.

Lastly, the researchers investigated the change in collective efficacy as teams gained more experience together over the trials. Indeed, previous research has indicated that collective efficacy develops in a group through common experiences and similarity in Team Task Awareness (Katz-Navon & Erez, 2005). Members of a team more familiar with their tasks should feel more efficacious. Expecting the same to occur in the present experiment, the researchers hypothesized that:

> H15: Collective efficacy will increase across Trials 1 through 4.

Participants

Participants (*N* = 111, making up 37 teams) were recruited using mailing lists of all staff and students at Iowa State University. Nearly every participant (89%, *n* = 99) reported working in teams at least once a month, and the majority (88%, n = 98) reported enjoying teamwork. Seventy-two participants (65%) reported playing videogames; just over half of those video games involved teams or cooperative play, on average (M=55%, SD =30%).

The participants completed the experiment in teams of three (n_{teams} = 37); nineteen (19) teams received public feedback, and 18 teams received private feedback. To ensure maximum timeslot usage, each participant, after consenting, was instructed to sign up for a time slot that already contained at least one person in them if possible before filling other time slots. Additionally, researchers alerted participants if they needed to reschedule due to an incomplete time slot. When participants arrived at the lab, they were welcomed to the study and key points of the informed consent were reiterated to them.

Procedure

When participants arrived, they were randomly assigned roles. Each team consisted of two spotters whose primary duties were transferring potential OPFOR, or opposing forces, to each other and the third member of their team, the sniper, at zone borders. The spotters used the keyboard to indicate how many OPFOR crossed at each of the zone borders. The sniper used the keyboard to indicate what level of threat each potential OPFOR posed.

The participants began the first of four trials after they watched a tutorial video and confirmed they understood the study. Each task lasted for five minutes, and each experimental session consisted of four trials. Participants received feedback on their performance, provided to them in real-time by an intelligent team tutoring system (ITTS). Throughout the trials, the feedback was either privately displayed only to the player to which it applied (heretofore referred to as “private feedback”) or was broadcast to the entire team (“public feedback”). In the Public feedback condition, participants would have normally received more than twice the feedback as in the Private condition, since each teammate would have received feedback triggered by all three team members rather than only the feedback triggered by his or her own actions. To balance the amount of feedback received in the Public condition, feedback was not given for Transfer events. Additionally, each team was randomly assigned to either a Spotter 1 – Sniper switch or a Spotter 2 – Sniper switch in the fourth trial. Just before Trial 4, all players were given a chance to ask questions about their role.

After each trial, the participants were asked to complete a post-trial survey. After the entire experiment, the participants were asked to complete a post-experimental survey and participate in an open forum discussion, led by an experimenter, with their teammates regarding the experimental environment and the feedback.

Participants were monetarily compensated for their time and payment was not contingent on performance. There was not a separate measure to filter for participants who were not engaged with the task, but the presence of experimenters in each of the separate rooms encouraged active participation.

Independent/Explanatory Variables

Feedback privacy. For each experimental session (all four trials), feedback privacy was manipulated to either be presented as Private, only to the person whose actions directly triggered it, or Public, all feedback was shown to everyone on the team.

Role naïveté. A direct consequence of the trial of the experiment and the introduction of a role switch in the final trial, role naïveté is an independent variable which relates to task and role experience. In the first trial, all participants were nascent in their roles, only having watched a training video before jumping into the task. In the second and third trials, the participants had experienced their present role once (in Trial 2) or twice (in Trial 3). In the fourth trial, either Spotter 1 or Spotter 2 (randomly assigned per experimental session) switched roles with the Sniper. Thus, the Spotter and the Sniper returned to role naïveté, albeit with their observations and the previous training video to guide them. Therefore, role naïveté serves as a second, within-subjects independent variable with three levels: (1) total naïveté (Trial 1), (2) experienced (Trials 2 and 3), and (3) partial naïveté (Trial 4 for those who swapped roles). Preliminary analyses of learning effects within this task showed that no learning occurred after the second trial (Ouverson, et al., 2018); thus, Trials 2 and 3 were deemed appropriate to collapse into a single role-experience category.

This concept additionally surfaces when the researchers refer to primary and secondary roles. Primary roles are the roles which participants are assigned at the start of the experiment. If a participant switches roles, they switch into their secondary role. One Spotter in each experimental session was randomly selected for the role switch in the fourth trial.

Video game and cooperative play experience. In a survey given prior to the start of the study, each participant's video game play frequency and multiplayer to single-player percentage was self-reported. While this was not a manipulated independent variable, this variable serves as a quasi-independent variable with three levels High, Moderate and Low frequency. For self-reported video game frequency, options were "Daily," "2-3 times a week," "Once a week," "2-3 times a month," "Once a month," or "Less than once a month." While this collapse reduced the nuance of the self-selected categories, it was meant to mirror the literature on the effects of video games, which has found daily, or nearly daily, interaction with certain video games to increase, for example, cognitive skills (Castel, Pratt, & Drummond, 2005; Strobach & Schubert, 2013). These options were collapsed into the three categories as shown in Table 4, the distributions of which are highlighted in Figure 5.

Table 4. Video game experience frequency coding choices.

Categories	Response Options		
High experience	*Daily*		
Moderate experience	*2-3 times a week*	*Once a week*	
Low experience	*2-3 times a month*	*Once a month*	*Less than once a month*

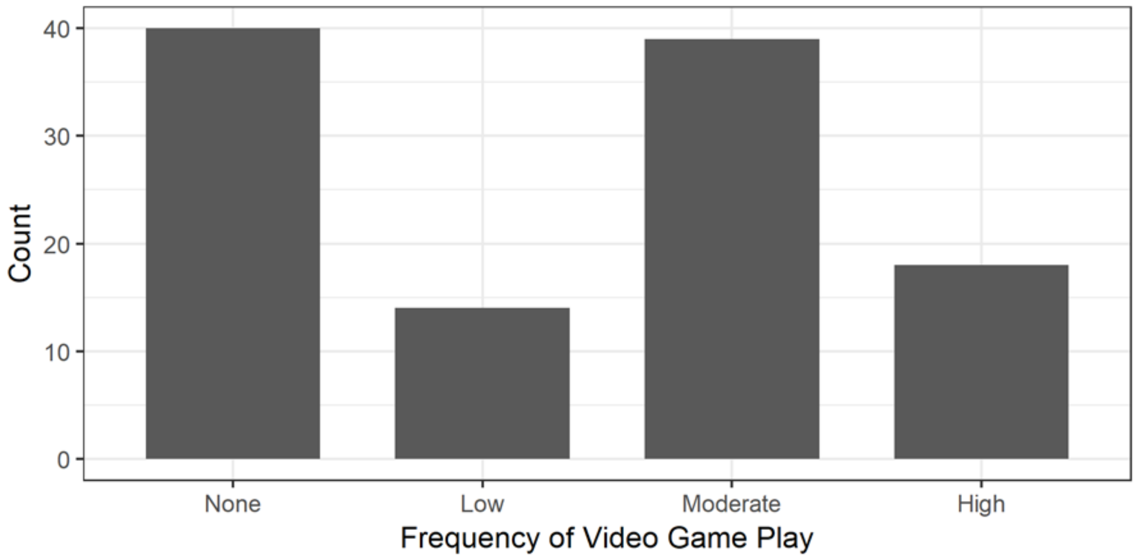

Figure 5. Distribution of video game play frequency among participants (n = 111).

The cooperative play distinction was collected as a self-reported percentage of multiplayer games played out of total gameplay time. The data were also split into three frequency levels based on the distribution – None included all non-responses and zero percent, Low frequency captured the values between one and 65 percent, and High included 65 to 100 percent. The distribution of these scores is highlighted in Figure 6. By examining this distribution, one can see that the distribution is not normal and there are three modes, one at zero, one at roughly 45, and one at roughly 85. This was interpreted as three distinct groups.

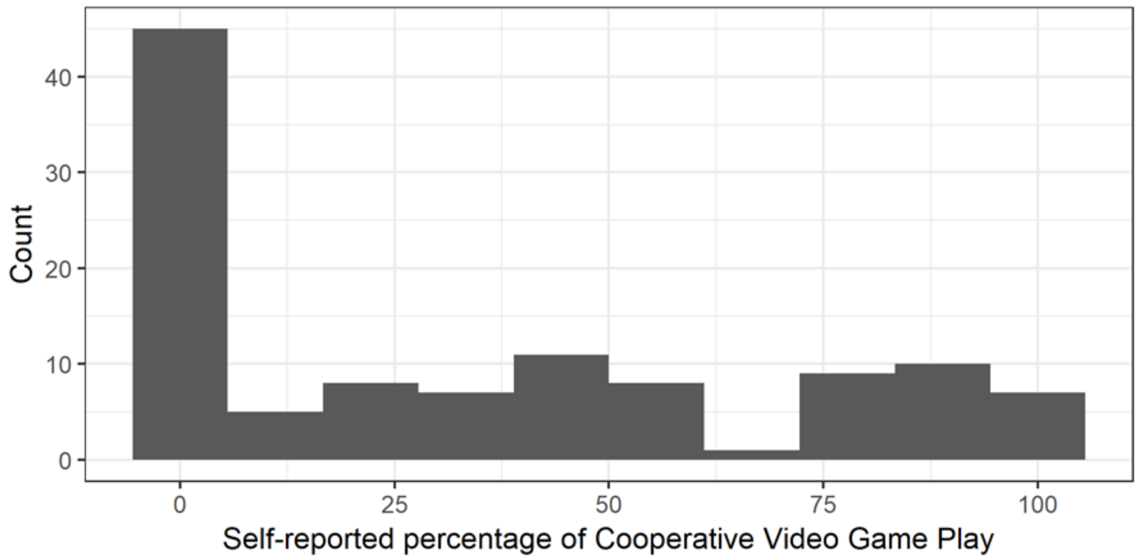

Figure 6. Distribution of cooperative video game experience among participants (n = 111). This distribution was split into three groups, one for all 0% and non-responses, one from 1% to 65%, and one for all values above 65%.

Teamwork experience. In a survey given before the start of the study, each participant's teamwork experience frequency was self-reported. While this was not a manipulated independent variable, this variable serves as a quasi-independent variable with two levels, High and Low frequency. For self-reported teamwork experience frequency, options were "Daily," "2-3 times a week," "Once a week," "2-3 times a month," "Once a month," or "Less than once a month." These options were collapsed into the three categories as shown in Table 5. This split was intended to keep the Low and High levels roughly even to facilitate comparisons between the two groups. The distributions of the participants per category are highlighted in Figure 7.

Table 5. Teamwork experience frequency coding choices.

Categories	Response Options		
High frequency	*Daily*		
Moderate frequency	*2-3 times a week*	*Once a week*	
Low frequency	*2-3 times a month*	*Once a month*	*Less than once a month*

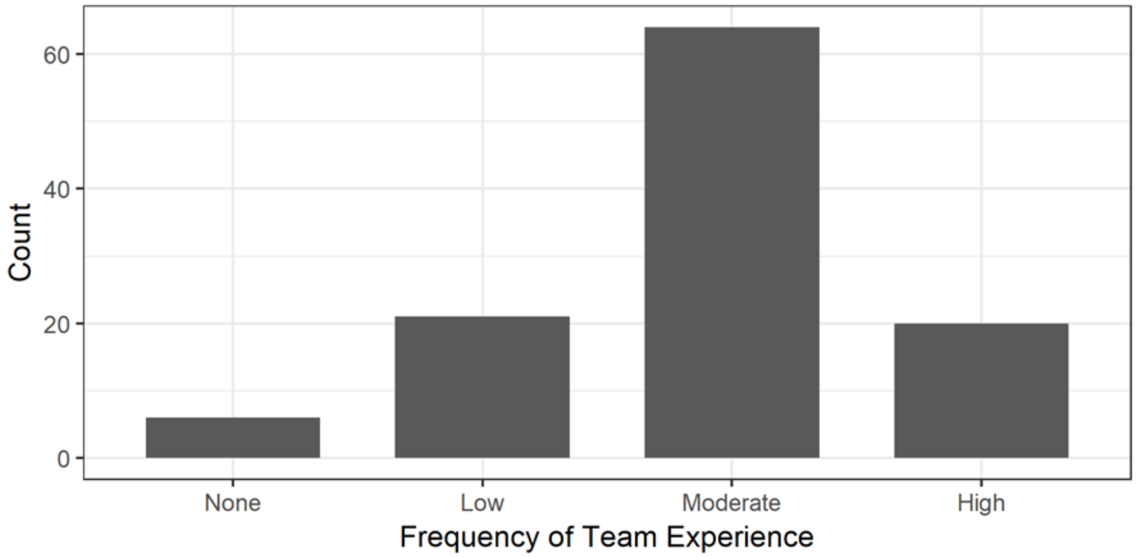

Figure 7. Distribution of the frequency level of previous team experience among participants (n = 111).

Teammate Familiarity. Participants assigned their own teams thus varying levels of prior teammate familiarity were observed. Just prior to the start of the study, a survey assessing baseline relationships within the teammates was given to each participant. For each teammate, the survey asked, “Have you met teammate X?” and the answers to these were recoded to 0 for “No” and 0.5 for “Yes.” Each participant could have a familiarity score of zero to one, and each team could have a familiarity of zero to three. These data were analyzed so that three levels of familiarity were established, as displayed in Table 6; the distributions of familiarity within teams is shown in Figure 8.

Table 6. Teammate familiarity coding choices.

Categories	**Scores (Based on response choices)**	
	Individual	Team
No familiarity	0	0
Partial familiarity	0.5	0.5 – 2.5
Full familiarity	1	3

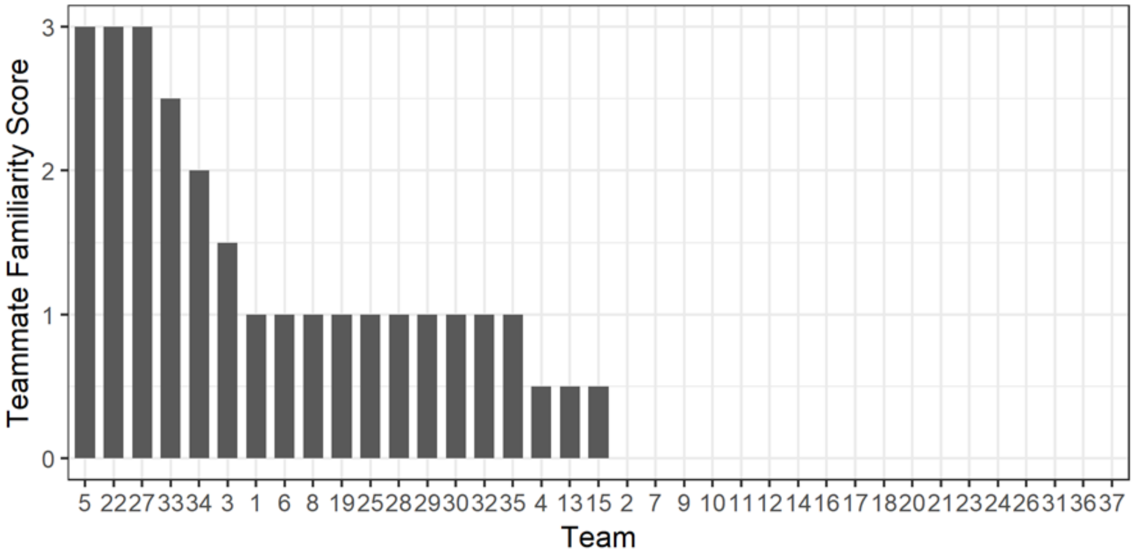

Figure 8. Distribution of team-level teammate familiarity across the team sample (n = 37).

Feedback helpfulness. A post-trial survey was given after each trial that asked if the participant noticed feedback ("Yes" or "No") and if the feedback was helpful (choices: "I ignored the feedback," "No, it was actually distracting," "No, it was not very helpful," "Yes, it was somewhat helpful," "Yes, it was very helpful"). Negative responses to the first question, "Did you notice any feedback during the task?" were included in the subsequent question as additional "I ignored the feedback" responses. Again, this variable was used as a quasi-independent variable in analyses. When comparing across trials, this variable was coded as shown in Table 7; the answer distribution is highlighted in Figure 9. No participants ignored feedback in all trials, so none were excluded from the analysis on this basis.

Table 7. Feedback use and helpfulness coding choices.

Categories	Response Options
0	*I ignored the feedback*
1	*No, it was actually distracting*
2	*No, it was not very helpful*
3	*Yes, it was somewhat helpful*
4	*Yes, it was very helpful*

Trial. There were four trials per experimental session. This introduced a time variable used to examine how experience influenced the dependent variables included below.

Dependent/Response Variables

Seven dependent variables, or response variables (the terms are used interchangeably in this thesis), are present in this study. The first four are derived from self-report, while the final three are derived from an analysis of player events as recorded by the GIFT Event

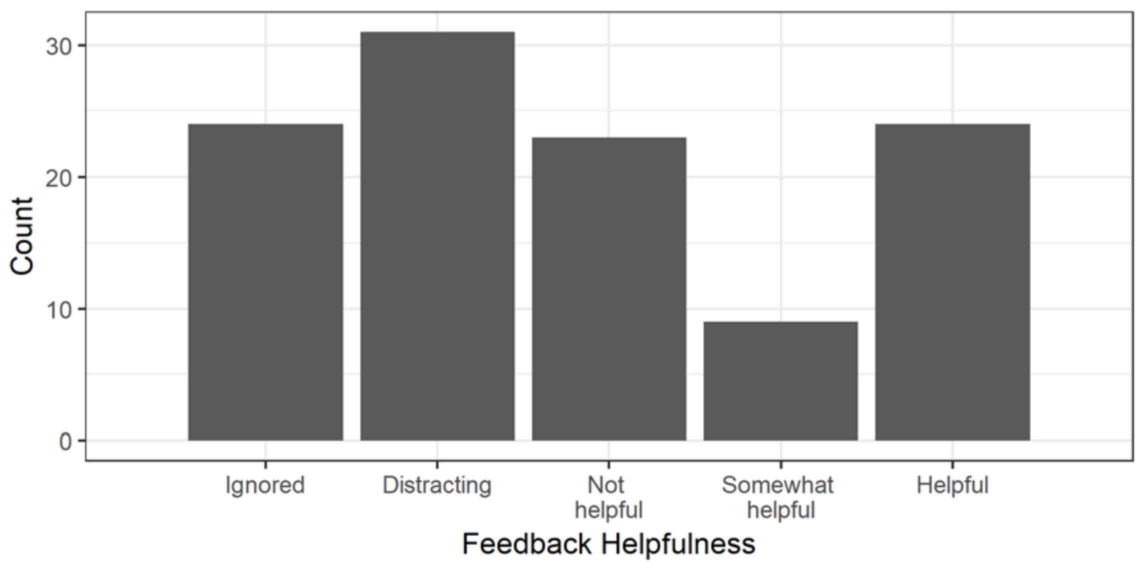

Figure 9. Distribution of answers to a question after each trial asking participants (n = 111) if they noticed and used the feedback. Data shown across all four trials.

Reporting Tool (ERT). Each of these is listed in Table 8 along with the metrics used to collect each of them, and each is described in more detail following the table.

Table 8. All dependent variables used in the present study and the metrics comprising them.

Dependent Variable	Metric	Frequency
Sniper Goal Awareness	Proportion correct on quiz of sniper goals (0 to 1)	Post-experiment (1x)
Shared Role Awareness	Similarity to teammates on Sniper and Spotter Goal Awareness Measures (0 to1)	Post-experiment (1x)
Team Task Awareness	Spearman rank correlation of quiz answers to correct answers (0 to1)	Post-experiment (1x)
Collective Efficacy	Average of five-question five-point Likert measure responses (1 to 5)	Each trial (4x)
Team Performance	Average percentage of the percent-correct of strings of Transfer-Acknowledge-Identify-Acknowledge-Assess actions (0 to 1)	Collected during each trial (4x), calculated post-experiment
Individual Performance	Average of the counts of correct responses for four separate actions: Transfers, Acknowledges (Sniper and Spotter), Identifies, and Assesses (0 to 1)	Collected during each trial (4x), calculated post-experiment
Communication	Percentage of prompt acknowledges (0 to 1)	Each trial (4x)

Shared Situational Awareness

Shared Situational Awareness (SA) was measured using quizzes given after the experiment to assess team members' understanding of the task and of their roles. Quiz results were analyzed for two measures: 1) how similar team members' mental models were to one-another's (Shared Role Awareness) and 2) how well participants' answers aligned with the tutorial information they were given (Sniper Goal Awareness and, separately, Team Task Awareness). This dual coding was used to examine the similarity of the teams' shared SA and the accuracy of their shared SA, respectively.

Sniper Goal Awareness. To derive this dependent variable, focused on the degree to which team members understood the Sniper's role, the quiz about goals of the Sniper (Table 9; Appendix D) was scored against a key based on the tutorial video. Originally, this measure was going to also include the Spotter goal answers, but after an analysis of the internal consistency, item difficulty and item discrimination, the Spotter quiz was excluded from further analysis. To calculate the final score, individual answers were compared to the answer key and ranked as correct (1) or incorrect (0). These binary scores were summed and divided by the total possible score of nine to create the individual Sniper Goal Awareness score with a possible range of zero to one.

Table 9. Quiz items and correct answers for the Sniper role's intended actions given at the end of the experiment. The scores on these items comprise the Sniper Goal Awareness score.

What are the Goals of the Sniper in this Task?	
☐ To identify targets new to their zone	☐ To keep count of how many targets have left and entered their zone
☐ To identify targets leaving their zone	☐ To keep count of how many OPFOR are on the map
☑ To assess the treats posed by targets	☐ To keep count of how many civilians are on the map
☑ To acknowledge what their teammates say	☐ To count the number of OPFOR wearing vests

Shared Role Awareness. To derive this dependent variable, each team member's quiz about the goals of the Spotter and Sniper roles (Table 10; Appendix D) was scored against the other members' answers. Because the Spotter and Sniper scores were statistically significantly correlated ($r(111) = .31$, $p = .001$), the two scales were collapsed into one measure: Shared Role Awareness. This ordinal variable included three levels: selection matches one teammate (1), selection matches both teammates (2), or selection matches neither teammate (0). For each participant, these scores were summed and divided by the total possible score of 18 to create the individual Shared Role Awareness score, which had a possible range of zero to one.

Table 10. Items from quizzes given after the experiment which were scored compared to teammates' answers to derive Shared Role Awareness.

What are the Goals of the Sniper in this Task?	
□ To identify targets new to their zone	□ To keep count of how many targets have left and entered their zone
□ To identify targets leaving their zone	□ To keep count of how many OPFOR are on the map
□ To assess the treats posed by targets	□ To keep count of how many civilians are on the map
□ To acknowledge what their teammates say	□ To count the number of OPFOR wearing vests

What are the Goals of the Spotters in this Task?	
□ To assess the treats posed by targets	□ To keep count of how many civilians are on the map
□ To keep count of how many targets have left and entered their zone	□ To count the number of OPFOR wearing vests
□ To keep count of how many OPFOR are on the map	

Team Task Awareness. To derive this dependent variable, each participant's response to the task-order quiz (Table 11; Appendix D) was Spearman rank-order correlated with the answer key. Distractors shared the lowest available rank. Therefore, possible scores on Team Task Awareness were from zero to one.

Table 11. List of steps to the task and the order in which they should occur, as given in the Team Task Awareness quiz. When a task step is erroneous, "NA" fills the order column.

Task steps	Order
Spotter 1 sees a target approaching the 1 pole	1
Spotter 1 transfers a target by pressing the 1 key	2
Spotter 1 transfers a target by pressing the E key	NA
Spotter 2 acknowledges his/her teammate's communication by pressing the E key	3
Spotter 2 acknowledges his/her teammate's communication by pressing the 1 key	NA
Spotter 2 sees a target near by the 1 pole	4
Spotter 2 identifies that a target has entered his/her zone by pressing the SPACEBAR key	5
Spotter 2 identifies that a target has entered his/her zone by pressing the E key	NA
Spotter 2 identifies that a target has entered his/her zone by pressing the 1 key	NA

Table 11. (continued)

Task steps	Order
Spotter 2 informs Sniper that a target has entered his/her zone	6
Sniper acknowledges his/her teammate's communication by pressing the E key	7
Sniper searches for a target in Spotter 2's zone in the direction of the 1 pole	8
Sniper spots a target and assesses the threat posed by the target	9
Sniper believes target to be a civilian and presses the C key	NA
Sniper believes target to be a civilian and presses the X key	NA
Sniper believes target to be a civilian and presses the Z key	10

Collective efficacy. Collective efficacy was collected prior to the start of the experiment and after each trial in the experiment. While self-efficacy is a measure of how well an individual thinks she or he can do tasks, collective efficacy is a measure of how well an individual thinks she or he can perform in a team. The data were collected on a 1 to 5 Likert-type scale with the anchors "Not at all confident" to "Extremely confident." Scale items are shown in Table 12.

Table 12. Scale items for the collective efficacy quiz used in this work.

	Scale items
1	My team can communicate important details in a timely manner
2	My team can satisfactorily communicate about important events
3	My team can accurately assess how to handle information we receive
4	My team can quickly assess how to handle information we receive
5	My team can accurately transfer information to one another

Team performance. Team performance is based on the percentage of OPFOR and civilians for which a team correctly executed the appropriate Transfers, Identifies, Acknowledges, and Assessments, across trials. The measure was calculated a priori using the data which was used by the tutor to decide when to give feedback and which message to present (and to whom, depending on the feedback condition). These data were compared to the known timings of simulation events – such as the crossing of an OPFOR between zones –

to find the percentage correct. The scores are percentages of the correct actions in the overall sequence. If a team member missed an action in the sequence, the team could still get partial credit.

Individual performance. Individual performance is a count of the missed Transfers, Identifies, Acknowledges, and Assessments of each participant. In this way, individual performance is not a metric that can be represented as a single number. Similar to team performance, this measure was calculated a priori using the same data. Instead of comparing these counts to the known timings of simulation events to create a percentage, the count of simulation events which did not have a matching player event was recorded.

Communication. Communication is operationalized as the percent of correct Acknowledge actions. This was chosen as the measure of communication because the Acknowledge action is used to inform a teammate that they have been heard, and even without these, the task would continue, more or less. In this way, the Acknowledge action is not primary to task completion, however, it does still add information to the team about member cognitive states in that it indicates to the team when communication is received. Acknowledge was analyzed as a count and as a percentage to account for the variability in required Acknowledge actions, since the number of necessary acknowledges depended on the number of Transfer and Identify actions of one's teammates.

Algorithm Verification

Communication and coordination, which are defined by (Salas et al., 2015), were used by the tutor to analyze team and individual performance and to give feedback. In order to use the same data for the analysis of performance dependent on explanatory or independent variables, the data needed to be cleaned, giving counts of player actions, and then re-evaluated so as to generate measures of performance. One of the difficulties inherent with the SwS task was that when many OPFOR were moving across a zone, and a Spotter said “Transfer,” it was not certain during post-experiment analysis which OPFOR the Spotter intended. Similarly, if a Spotter receiving the OPFOR noted “Identify” three times for three OPFOR that crossed, it was not clear to the data analyst which of the OPFOR should be mapped to which of the Identifies.

To resolve this data analysis ambiguity, data from the GIFT Event Reporter Tool (ERT) were visualized such that each learner event, or input from the player, was represented as a vertical line, and each simulation event, or scripted game event, was represented as a color-coded horizontal bar, as shown in Figure 10. In the visualizations, pink player-event lines symbolized Transfers, light-green dash-dot lines were for Identifies, and blue dashed lines represented Acknowledgments. For the simulation events, the colors roughly represented the zones before and after an OPFOR (of which there are 40) crossed from one player’s territory to the other (the small vertical black line). Zones at left of the crossing point

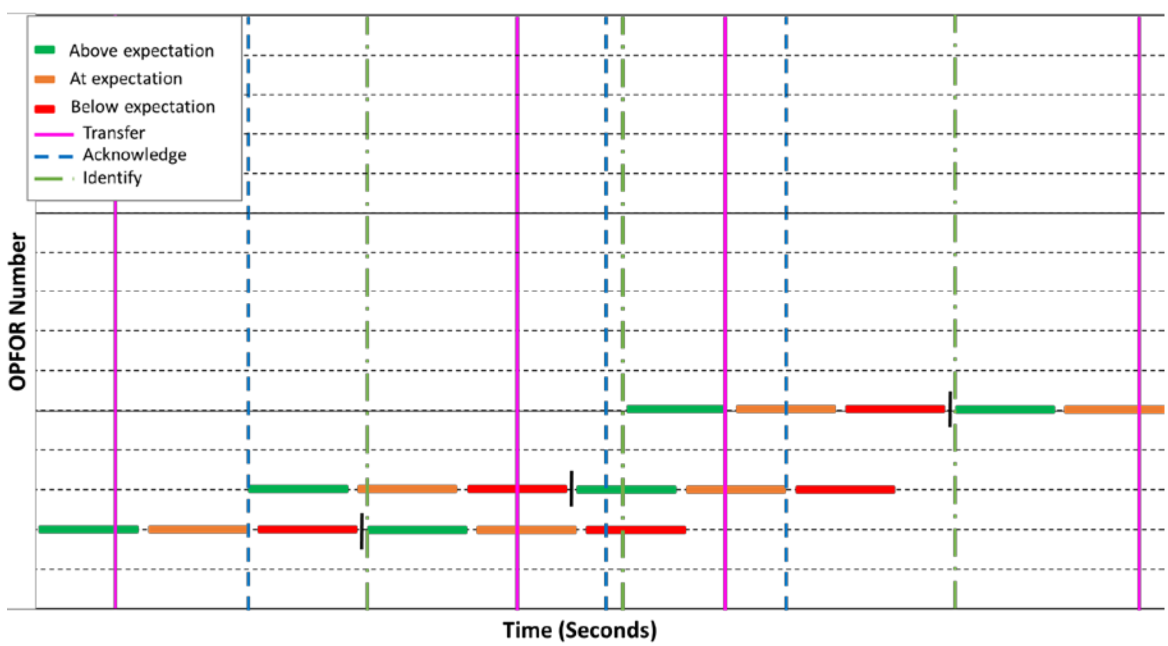

Figure 10. Example timeline of a subset of OPFOR crossings (horizontal thin lines) and player events (vertical lines), data which are parsed from the tutor's event log.

represent the ideal and less than ideal timing for a spotter to Transfer an OPFOR, while zones to the right of the crossing point correspond to the ideal and less than ideal timing for a spotter to Identify an OPFOR. The first zone is “above expectations,” second is “at expectations,” and last is “below expectations.”

Because of the high volume of non-specific player events, the tutor’s expectations of the player’s performance could be misleading when in a high OPFOR-traffic period of the simulation. In other words, a participant may have been transferring three OPFOR at appropriate times, but the tutor’s algorithm may have counted the first one as having been too early or the last as too late due to delays elsewhere in the team. Therefore, the researchers decided to use human coding during the first pass at analyzing the performance of teams and participants. For this reason, a subset of these visualizations was evaluated by three independent judges, who then compared their evaluations and processes to develop a single method with robust interrater reliability (Gilbert et al., 2016). That method was followed for the categorization of events in the remainder of the visualizations.

This original human rating method was later replaced, however, with a matching algorithm based on the original values the tutor used to assess performance in real time. Although participants were never given their “performance metrics,” since feedback was given based on these assessments, numbers pulled in this fashion were determined to be a better representation of the performance values to which participants compared themselves. More details about this algorithm are described in Ostrander (2019).

Upon completion of the evaluation study for SwS, the algorithm was modified to include information about the sniper role (acknowledges, assessment occurrence percentage, and correct assessments). Before using the data generated from this algorithm, the author verified its accuracy by conducting a visual comparison of the algorithm's ratings of missed events to the visualizations for three randomly-selected trials for random teams: nine observations in total. Once satisfied that the data were accurately reported by the modified algorithm, the author used these data to evaluate individual and team-level performance.

Data Analysis Plan

Preliminary analyses. Normality of the sample on several dimensions was assessed using quantile-quantile plots of the residuals, while homogeneity of variances was assessed by examining scatterplots. These are featured in Appendix B, and while some of the models did violate the normality and homogeneity of variances assumptions, no adjustments were made to the models. Models were sufficiently normal and homogenous for these analyses.

To confirm the internal consistency of measures used in this research, Cronbach's alpha was calculated (Cronbach, 1951) for the collective efficacy scale, and Kuder-Richardson formula 20 (Thompson, 2010) was used to calculate the coefficient for the role understanding scales (of which there are two). KR-20 is interpreted in the same manner as Cronbach's alpha, but is more appropriate for dichotomous data. Both measures are generally considered to be adequate if the coefficient is greater than .70, although this cut-off is arbitrary (Cortina, 1993). The Spotter Role Understanding scale had an alpha value of -0.04 and was therefore excluded from analyses. The Sniper Role Understanding scale had an alpha value of 0.73, and the overall Shared Role Awareness scale had an alpha value of 0.71. The Collective Efficacy scale was found to have a Cronbach's alpha of 0.95. Cronbach's alpha was not calculated for the Team Task Awareness scale, which required participants to correctly order and sort the items, because the assumptions for the test of internal consistency were not met.

Experimental analyses. Hypothesis testing was done by fitting a linear mixed-effects model using the restricted maximum likelihood (REML) criterion (Bates, Maechler, Bolker, & Walker, 2015) and estimated marginal means (Lenth, 2019). This method was chosen to account for the dependence of the individual participants in the study on their teammates and team experience for performance measures, and to a certain extent, self-report measures like collective efficacy. Tukey's (Honest Significant Difference) HSD was used to account for multiple comparisons in pairwise differences. Cohen's *d* (Cohen, 1988) was supplied for each difference to delineate the effect sizes, combating Type I and II errors (Sullivan & Feinn, 2012). Cohen (1988) indicated that, when interpreting effect sizes, $d = 0.2$ showed a small effect, $d = 0.5$ could be considered a medium-sized effect, and $d = 0.8$ showed a large effect.

Summary

Now that the task and methods for the SwS study have been detailed, the following chapters discuss findings and the implications for those findings. Chapter 5 describes an analysis of the communication and self-efficacy variables, while Chapter 6 describes an analysis of the performance variables. The hypotheses tested will be summarized prior to reporting results.

CHAPTER 5. ANALYSIS OF COMMUNICATION, SHARED SITUATIONAL AWARENESS, AND FEEDBACK WITHIN A THREE-PERSON INTELLIGENT TEAM TUTORING SYSTEM

Based on a manuscript to be submitted to the journal *Computers in Human Behavior*

Author list:

Kaitlyn Ouverson[1]; Alec Ostrander[1]; Jamiahus Walton[1], MS; Adam Kohl[1], MS; Stephen Gilbert[1], PhD; Michael Dorneich[1], PhD; Eliot Winer[1], PhD; Anne Sinatra[2], PhD

[1] Iowa State University, Ames, IA; [2] Natick Soldier Research, Development, and Engineering Center (NSRDEC) Simulation Training and Technology Center (STTC), Orlando, FL

Kaitlyn Ouverson's role in this research included contributions to the experimental design and data collection and data coding, as well as sole responsibility over participant recruitment and data analysis, and primary authorship on the chapter/paper. She specified the parameters of the tutoring environment that would be needed to test hypotheses about the intelligent team tutoring system, coordinated the research assistants for data collection, and helped collect data, herself. Lastly, she introduced a method of measuring situational awareness in the experimental design and analyzed the data from those survey items.

Abstract

This work assesses how the privacy of feedback generated by an Intelligent Team Tutoring System (ITTS), teammate familiarity, role experience, and prior virtual and face-to-face team experience affected the team skills of shared situational awareness (SA) and communication. Previous work has focused on outcomes for task skills, with limited focus on team skill training. While other results offer privacy recommendations for team feedback, implementation within an ITTS has yet to be presented. Thirty-seven teams of three participants were given just-in-time private (individually-delivered) or public (team-delivered) performance feedback during four five-minute trials. In the fourth trial, two of the three participants switched roles. When participants were learning new roles, communication was found to be significantly lower. Feedback type, role switching, teamwork experience, or teammate familiarity had no statistically significant effect on communication or shared SA, but a potential marginal effect of shared SA depending on role switching was noted. Results, while not significant, delineate an approach for analyzing team data and point toward considerations for future work. Suggestions for overcoming study limitations are provided so that this work may serve as a platform for further research. This study establishes a foundation for future research on designing effective ITTSs that train interpersonal skills to nascent teams.

Keywords: situation awareness, team cognition, shared/team mental models, team training, intelligent tutoring system

Introduction

Intelligent Tutoring Systems (ITSs) in a variety of domains have demonstrated success by assessing learners' task skill mastery (Graesser et al., 2017; Koedinger, Brunskill, Baker, Mclaughlin, & Stamper, 2013; VanLehn, 2011). In the past decade, additional research has explored ITS learners' affect and motivation (Mumm & Mutlu, 2011; Price et al., 2018; Sabourin, Mott, & Lester, 2011; Yang & Dorneich, 2016), metacognitive factors not related directly to the task. More recently, efforts to create Intelligent Team Tutoring Systems (ITTSs) that offer automated coaching to teams expand this non-task focus to the challenge of how to tutor team skills, independent of a task. The present research evaluates the impact of an ITTS on the team skills of shared situational awareness and team communication.

Traditionally, teams are defined as a group of two or more members, each with specific tasks which require coordination of information and activities to reach some common goal or objective (Salas, Dickinson, Converse, & Tannenbaum, 1992). The coordination of tasks and information exchange are actions that may require some measure of training for a team to interact successfully; however, training which focuses on interpersonal skills has seen limited attention in intelligent tutoring system research (Lane et al., 2007; Orvis et al., 2010; Riggio & Lee, 2007). Team training which encompasses interpersonal skills, or team skills, in addition to task skills is becoming increasingly important as more teams interact primarily virtually. It follows that as teamwork continues to become virtual, dispersed team members should be able to interact effectively in virtual settings, established through distributed team training, and training should be possible via virtual means. In the present paper, the authors assess the impact of ITTS feedback, teammate familiarity, and virtual team experience on shared situational awareness (SA) and communication.

Background

Starting with systems that helped keep human trainers better apprised of the team's whole performance (Zachary et al., 1999), work investigating ITSs has branched from simple algorithms to social tutors giving team- and individual-level assessment (Kumar et al., 2010; Ostrander et al., 2019; Walton et al., 2015). Within the environments of ITTSs, behavioral markers are used to identify team metrics (Salas et al., 2007; Sottilare et al., 2018). From these metrics, ITTSs use feedback to present information about team and individual performance (Ostrander et al., 2019; Sinatra et al., 2018; Walton et al., 2018). Feedback from the tutor, in turn, influences teammates' actions, interactions, and shared SA. The present paper explores the impact of feedback, experience, and familiarity on communication and shared SA within team training using an ITTS. Each of these topics is explored further below.

Teams and Team Training

Teams have been a topic of study for nearly a century, starting with examinations of groups working together in factories and developing into a depth of work seeking to uncover the components which make up a good team (Bisbey et al., 2019). Early work focused on understanding group work, focusing on how to make workers more efficient in their individual roles and thereby increasing the group's performance. However, that work was not the interdependent team work that is studied today. Starting in the 1980's researchers began to focus on the need to understand "teamwork" (Dyer, 1984, Hritz et al., 1983, as cited in Bisbey et al., 2019). The outcomes of specific team failures (such as the incident of the USS *Vincennes* shooting down a civilian flight) largely influenced this push, and team training initiatives, such as Team Dimensional Training (TDT) and Crew Resource Management (CRM) training, were developed to ensure failures were mitigated.

Previous Intelligent Team Tutoring Systems (ITTSs)

Since the late 1990s, ITTSs have been developed to facilitate training on team tasks from Naval air defense training (Zachary et al., 1999), mechanical engineering (Kumar et al., 2010), group shopping (Walton et al., 2015), collaborative problem solving (Fiore et al., 2017), and team-coordinated surveillance (Ostrander et al., 2019). One of the first ITTS-like systems, the Advanced Embedded Training System (AETS), facilitated Naval air defense training (Zachary et al., 1999). The AETS monitored the learners' button presses, speech, and eye movements to supplement the work of a human trainer. The human trainer's time focused on aggregating data from the AETS into a team-level after-action review, while automated task feedback was given just-in-time to individuals by the AETS.

In the AETS, team members were assigned specific jobs, and feedback on performance was given by both the software agent and the human trainer. In the Team Multiple Errands Task (TMET; Walton et al., 2015) the software agent, or tutor, supplied real-time individual and team-level feedback to a team of three as they completed a multiplayer virtual shopping task. The TMET extended a classic single-person shopping-based cognitive task to a team of three.

The team member roles required by a team task often play an important role in the dynamics of the team. In TMET, the team members' roles were homogeneous, with no specific job roles or background training. In education and the workforce, team members often play different roles. Software development teams, for example, may consist of designers, engineers, and user researchers who work together to ship new products. While homogeneity makes the study of a team simpler and more controllable, the ecological validity of the team tutor in such situations is decreased.

While the TMET tutor was not embodied or personified, and all feedback was given as brief phrases or data visualizations based on individual performance or team scores, a different kind of team tutor, Avis (Kumar et al., 2010), was able to give feedback to a team through conversational dialogue, acting as a guide for learning underlying concepts of mechanical engineering. While Avis could be considered an ITTS, the tutor did not provide feedback for the team as a whole; it instead attended to each learner's conceptual understanding. Without team-level feedback, team skills are not actively trained, and therefore, Avis can be referred to as a socially capable tutor. Further, the use of teams and conversation by Avis was used to facilitate the learning of the material, rather than the improvement of team skills.

The agents developed within the Programme for International Student Assessment (PISA) 2015 (Fiore et al., 2017) engaged learners in conversation in much the same way as Avis. Instead of tutoring as a facilitator, the agent (and sometimes multiple agents with various skills) worked collaboratively with the learner as a peer to solve problems. In this way, the PISA 2015 agents are not tutors, although their goals are similar to agents in tutoring roles. Additionally, tasks always involved only one human teammate, rather than a multiple-human team. While the PISA 2015 encouraged the use of soft skills in engaging conversations, the inclusion of only one human per team limits their applicability to teams of two or more. As was true for TMET, this limits the ecological validity of any points made regarding team performance since no human-human coordination was necessary for the completion of the tasks.

Unlike Avis and the PISA system, the Surveillance Team Tutor (STT; as well as AETS and TMET) contained a fast-paced, high-cognitive load psychomotor performance task that required steady focused attention and did not typically allow for conversational dialogue (Ostrander et al., 2019). This tutor was designed to train dyadic teams on a military task using just-in-time feedback which was tailored to individual player actions. The STT offered two types of feedback to its users, which were referred to as Team feedback and Individual feedback.

The statement of “Team” or “Individual” feedback is not as specific as it first seems. Does Team feedback reach everyone on the team, or is it developed based on the team’s cumulative performance? Or, is it called Team feedback because it never specifies the audience for its statements? To account for this opacity, the researchers reference the three “dimensions of feedback,” as shown in Figure 11. By referencing the configuration of the feedback on these three dimensions, the specificity of the manipulation is clarified. In the case of the STT, Team feedback followed a TTT configuration, where assessment, delivery, and address all occurred at a team level, while Individual feedback followed an III configuration (Ostrander et al., 2019). These two were chosen because they represent the two most different configurations for feedback in this three-dimensional space of feedback characteristics. In the STT, feedback was coordinated using the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg, & Holden, 2012). Participants in a control condition were given no feedback.

The military task for STT was developed in Virtual Battlespace 2, a serious game engine. In the task, two spotters were positioned on top of a building in the middle of a virtual village environment which included walls as obstacles between which OPFOR (OPposing FORces) could run. Each spotter was responsible for watching a zone consisting of half of the environment and alerting his or her teammate to OPFOR who neared that

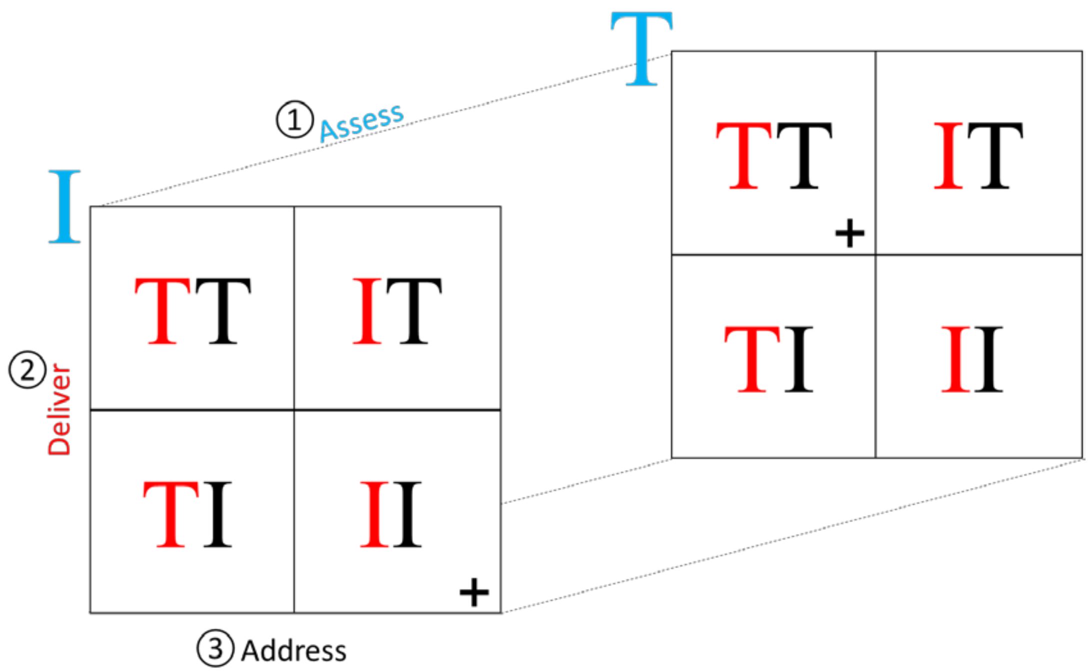

Figure 11. *Three dimensions of feedback considerations: team vs. individual basis for 1 Assessment, 2 Delivery, 3 Address. The + identifies comparisons made in previous research by the authors (Bonner et al., 2017; Ostrander et al., 2019; Sinatra et al., 2018).*

teammate's zone. The full task consisted of a Transfer event, in which one spotter alerted the other to an approaching OPFOR; an Acknowledge event, in which the receiving spotter acknowledged the transfer; and an Identify event, in which the second spotter noted receiving the OPFOR into his or her zone. Teammates passed this information to each other via a verbal communication channel and to the tutor using pre-assigned keyboard keys.

While the tutor was shown to have limited impact on the performance of participants and their teams (Ostrander et al., 2019), there were promising results related to the impact of feedback on shared mental models, and overestimation of performance. Feedback on the Acknowledge subtask did result in fewer errors, and Team feedback, in general, reduced the tendency of individuals to rate their teammates as having performed poorly. Lastly, in the Individual and No feedback conditions, participants' self-ratings of individual performance did not correlate significantly with their tutor-assessed performance, while they did in the Team feedback condition. These results suggest promise for team training with an ITTS.

Teamwork Ability: Communication, Cognition, Shared SA, and Familiarity

In order to train a team, metrics for team success and teamwork ability must be identified. One such metric for a distributed team may lie in the team’s ability to leverage benefits and minimize deficits of computer-mediated communication (CMC) technology (Alsharo et al., 2017; Pinjani & Palvia, 2013). Communication is clearly central to distributed teams and is recognized as a promising marker for team ability in these situations. This is echoed in previous research, as information sharing is a recognized path to better team outcomes like task success and creative solutions (Alsharo et al., 2017).

Communication which facilitates the flow of information decreases situational invisibility and also raises teammate shared situational awareness (Cramton et al., 2007), which is a team-wide understanding of the shared goals and the tasks required to achieve them, as well as an awareness of each member’s environment. In teams, situational-communication must occur between team members; however, in a training scenario, the supervising agent or person could inject such information.

The team’s shared situational awareness influences attribution, or how teammates assign responsibility for actions, whether those actions are attributed to the teammate’s personality or an external circumstance (Cramton et al., 2007). In addition, role switching, which establishes a grounded understanding of the team’s roles, can create a more functional environment in which teammates are able to anticipate member actions and cover extraneous responsibilities when necessary (Sottilare et al., 2011).

Teamwork ability is often quantified using conversation-based team metrics. However, researchers have identified measures of teamwork beyond conversation-based metrics (Salas et al., 2009; Wiese et al., 2015) that can help identify teamwork abilities in vivo, where outcomes are not directly quantifiable or occur over an extended time period, and synthesized research on team success metrics in an attempt to standardize the terminology and direct future research. Of particular importance to the current research effort is the concept of Cognition. Salas and colleagues (2015) detail the importance of team Cognition, which describes a group's ability to function cooperatively toward a common goal. To do so, all team members should be aware of the common team and shared situations – who must be at this position at what time if we are going to succeed? In this way, team Cognition is a superset of shared situational awareness.

The current paper investigates how the amount of role experience influences shared situational awareness (SA). Sætrevik and Eid (2014) note the importance of SA for each member of a team, stating that for a team to perform its best, each member must understand their tasks within the team. This understanding is developed through training and experience with the task. Therefore, it is hypothesized that:

> H1: For the participants who experience more than one role, shared situational awareness will be higher, compared to participants who experience only one role.

Beyond just increasing situational awareness and group cognition, communication is the cornerstone of group affinity (Nardi, 2005; Oren & Gilbert, 2011). This is especially true of communication which is not explicitly meant to increase task knowledge, for those head-nods and casual greetings are what contribute most to individuals' feelings of belongingness (Nardi, 2005). Early belonging, in turn, builds commitment to team goals, supporting trust in peers, performance and team satisfaction (Haines, 2014). While the casual conversations which build group affinity are natural for collocated teams, distributed teams may need some assistance in establishing and maintaining interpersonal bonds, especially if they are new at using CMC (Oren & Gilbert, 2010).

Another way to understand the impact of group affinity is via the familiarity of a team's members with one another. Friendship has been shown to improve performance (Mason & Clauset, 2013), but familiarity does not necessarily need to mean a relationship on the level of "friends." Familiarity on a professional level fosters an understanding of teammates' skillsets and strengths, thereby increasing the effectiveness of a team, especially under high workloads (Smith-Jentsch et al., 2015). Lastly, teammate familiarity offers an easy path to belongingness, which is an important component of team performance (Haines, 2014). An individual who knows at least one of his or her teammates should perform better on the team, communicating more with teammates. It is therefore hypothesized that:

> H2: Participants who are familiar with at least one teammate will have better communication than those who are less familiar.

Previous experience in teams across domains should logically impact an individual's future performance within a team. Research points toward the impacts of previous teamwork experience on future teamworking attitudes (Rudawska, 2017) and the understanding of how to perform well on future teams (Hirsch & Mckenna, 2008; Reagans et al., 2005; Rentsch et al., 1994). The tie between team experience and performance has, thus far, been indirect; an individual's experience has influenced his or her schema of teamwork, which in turn influences that person's performance in a team setting. Because of this, the researchers hypothesize that:

> H3: Persons who work in teams more often will have higher shared situational awareness.

Similar to experience with teamwork, experience with one's role is important to team performance. Reagans and colleagues (2005) found that experience within a team, and with one's role on that team, impacted team performance. Experience with a team also has been shown to influence the team's shared mental models – a component of shared SA – which have been shown to decrease the need for communication (Carpenter et al., 2008). However, the present study used feedback to actively encourage team communication, so it is expected that communication would be unaffected by the establishment of shared mental models.

In the fourth trial of the present study, two participants on each team switched roles. Role switching establishes a grounded understanding of the team's roles and has been shown to foster more effective communication long term (Sottilare et al., 2011). However, the short term effect of the role switch is expected to increase task load (as shown in Ouverson et al., 2018) and, therefore, communication errors. Therefore, it is hypothesized that:

> H4: When stepping into a new role for the first time, communication errors will be higher than when a person has experience in that role.

Feedback

Proper feedback, which is aimed at increasing awareness of the task or of the process, or at increasing self-regulation, has positive impacts on learning (Gilbert et al., 2017; Timperley & Hattie, 2007). By identifying the goals of each role within the team, in essence creating a model of an "expert team," feedback can be created for the task. Additional attention to the components of teamwork, such as those compiled by Salas et al. (2015), can be used to form the pedagogy to support learned team behaviors, such as communication. While performance on taskwork has been shown to be most receptive to privately-given feedback, teamwork (e.g., communication) has been responsive to publically-given feedback (Geister et al., 2006; Mumm & Mutlu, 2011), although results have been mixed (Peñarroja, Orengo, Zornoza, Sánchez, & Ripoll, 2015).

Also worthy of consideration is the way in which feedback alters the willingness of team members to welcome the help of an ITTS, as inappropriate etiquette (Dorneich et al., 2012; Walton et al., 2014) or excessive messages (Price et al., 2018) have been shown to harm learner performance in ITS situations. This can be solved through attention to the quantity of messages and the affect portrayed by those messages. For example, by reducing the number of messages a tutor supplies during expected times of high workload, the chance of the tutor interrupting learners is reduced, and thus etiquette norms are maintained.

Beyond the content of feedback – its pedagogy and affect – it is important to consider how to address and deliver feedback, and what behaviors to assess when crafting said feedback. In many ways, this is a matter of individual versus team-level feedback, again referencing the dimensions of feedback. Behavior assessment at the team level means that feedback is generated for team behaviors, like coordination and communication, while individual assessment measures individual tasks, like the speed of communication and accuracy of responses. Feedback delivery refers to the difference in the recipient of the feedback, either the individual for whom it is relevant or the whole team. The last consideration for feedback design is the audience to whom the feedback is addressed, either to the team or to each individual. Not all combinations of these considerations make pedagogical sense, for example, feedback which is gathered from an assessment of the Team, delivered to the Individual, but addressed to the Team. However, this framework for feedback offers a platform for researching optimal feedback characteristics in different team contexts.

The researchers expect that feedback which is delivered to the whole team, here referred to as "public," would have a positive effect on the shared SA since the use of a team-

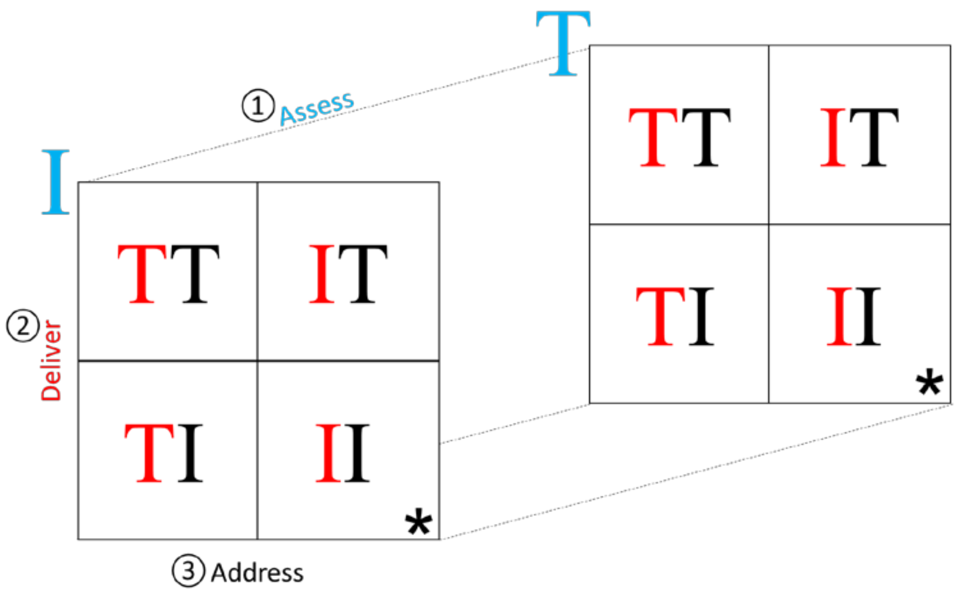

Figure 12. *The configuration of the three dimensions of feedback explored in the present study.*

level delivery feedback dimension will result in more shared information than individualized feedback (Geister et al., 2006). Additionally, public feedback (a TII configuration, as compared to private feedback, an III configuration; see Figure 12) would allow teammates to keep track of their whole-team communication performance, correcting and encouraging others to correct as necessary. It was hypothesized that:

> H5: Public feedback will result in higher shared situational awareness than private feedback.

> H6: Public feedback will result in lower communication errors.

Methods

Participants

Participants (*N* = 111) self-identified as 45 females, 61 males, and five persons who did not self-identify as either gender or preferred not to disclose their genders. The average age of the sample was 23.2 years of age (SD = 7.8). Nearly every participant (89%, *n* = 99) reported working in teams at least once a month, and the majority (88%, *n* = 98) reported enjoying teamwork. Seventy-two participants (65%) reported playing videogames; just over half of those video games involved teams or cooperative play, on average (M=55%, SD =30%).

The participants completed the experiment in teams of three (n_{teams} = 37), which were determined by experiment sign-up selection, which was mostly random but did allow for groups of friends to participate together. As such, 36% of participants (*n* = 40) had met at least one person on their team before the experiment.

Procedures

Participants were recruited using an all-student and staff mailing list at a large Midwestern University. Before they signed up for a timeslot, participants were required to give informed consent and basic demographic information. When they arrived, participants were randomly assigned roles. Each participant completed a familiarity survey, which asked whether they knew their fellow participants and was instructed to watch a tutorial video. The video introduced the task, the environment, and the controls for each role (See Figure 13 for example tutorial screens). Participants entered separate rooms to use individual computers, but there was an open audio channel connecting the three rooms. Participants began the first of four five-minute trials after they confirmed they understood the study. In the fourth trial,

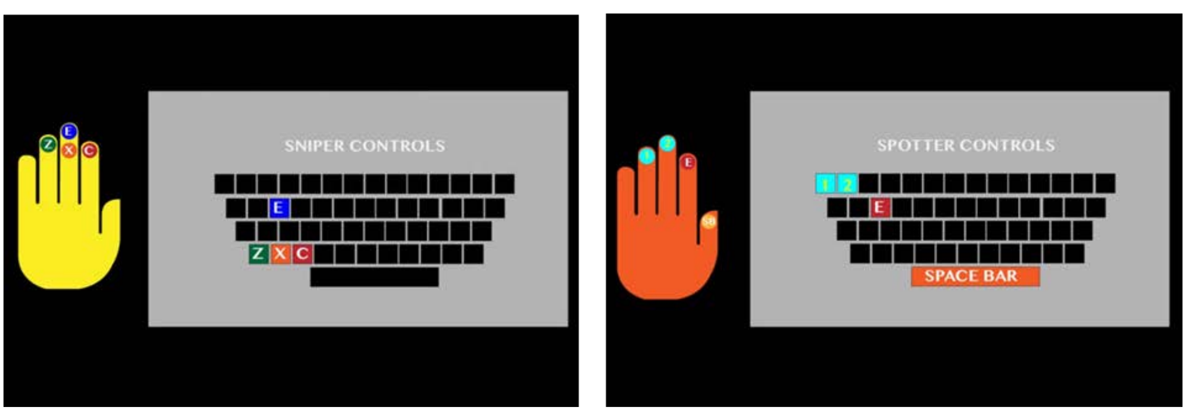

Figure 13. Sample tutorial video screens showing controls for each position and

two of the three teammates switched roles, and just before starting in the new configuration all players were given a chance to ask questions about their role.

After each trial, the participants were asked to complete a post-trial survey. After the entire experiment, the participants were asked to complete a post-experimental survey and participate in an open face-to-face discussion, led by an experimenter, with their teammates regarding the experimental environment and the feedback.

Task overview. The goal of this team task, the second version of the STT, was for a team of three to pass targets from one teammate to another as they moved from one side of the virtual environment to the other, ending with a threat assessment in which the level of potential threat posed by each target is reported to the tutor and to the team. Each team consisted of two spotters whose primary duties were transferring potential OPFOR, or opposing forces, to each other and the third member of their team, the sniper, at zone borders. The sniper used the keyboard to indicate what level of threat each potential OPFOR posed.

The full sequence of required subtasks in each trial is detailed in the following example:

Spotter 1 sees an entity in her zone heading towards the one-pole boundary. “One at pole one,” she asserts, pressing the assigned key (here “1”) transferring the entity to Spotter 2.

“Okay.” Spotter 2 strikes the “E” key, acknowledging the transfer. After the entity enters his zone, Spotter 2 alerts the Sniper, “There’s someone at pole 1, Sniper,” and presses the spacebar.

“Got it,” the Sniper acknowledges (again, using the “E” key), and using the “B” key to zoom in on the entity and assess the threat it poses. Seeing that the entity is a civilian, the

Sniper keys “Z,” rather than “X” or “C” which are used to signify an OPFOR wielding a gun but not wearing a vest and an OPFOR wearing a vest and/or wielding a gun, respectively.

Task feedback. In the first iteration of the STT, feedback derived from team assessment, delivered and addressed to the team (TTT configuration) was compared to that derived from individual assessment and delivered and addressed to the individual (III configuration). The SwS team tutor was designed to give feedback during the SwS task. To do so, each participant’s performance was tracked for each of the expected user events. The tutor used information about how users improved over time and what feedback and feedback triggers were present to coordinate ongoing performance feedback.

By consulting the programmed conditions by which that feedback was to be triggered or not triggered, the SwS tutor gave feedback for the Transfer, Acknowledge, Identify, and Assessment subtasks. The type of feedback (either Private or Public) was determined by the tutoring paradigm which was selected upon task set up. Subtasks were evaluated as either *Below Expectation*, *At Expectation*, or *Above Expectation* as specified in Figure 14. To avoid having feedback triggered too frequently (by every player action), a “bucket” system was used. When the requisite number of actions were documented at a single level of

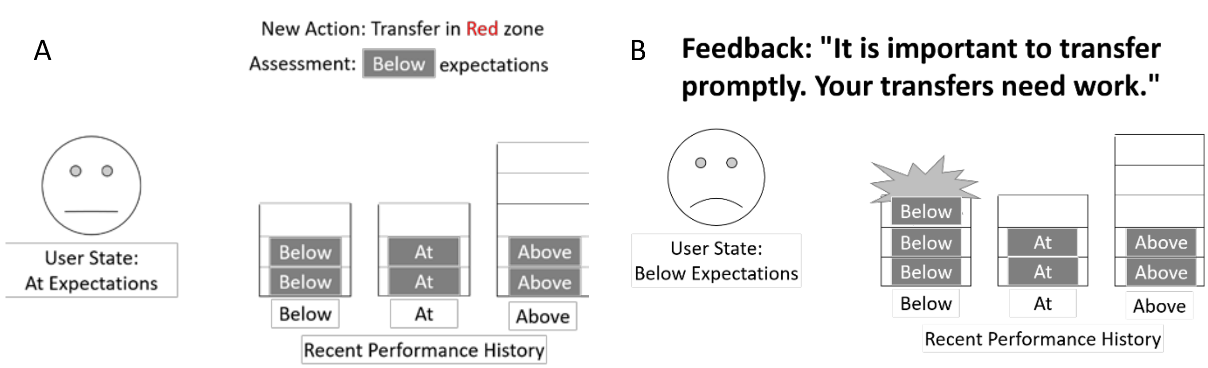

Figure 14. *Feedback was not given after ever player event. After a new user state was identified (B), players were given constructive feedback on their recent performance. Adapted from figure originally featured in Ostrander et al. (2019).*

performance, feedback was triggered, and the user state moved accordingly, if appropriate. This feedback trigger adjustment was further detailed in Gilbert et al (2017).

Examples of feedback given to the Spotters and Snipers are shown in Table 13. The Public feedback condition would have normally resulted in participants receiving more than twice the feedback as those in the Private condition, since each teammate would have received feedback triggered by all three team members rather than only the feedback triggered by his or her own actions. To balance the amount of feedback received in the Public condition, feedback was not given for Transfer events; since the Acknowledge actions are directly tied to the Transfer or Identify actions, this feedback was framed as being triggered by the Transfer-Acknowledge or Identify-Acknowledge pairs in the Public condition.

Table 13. Examples of feedback given for each player action in the task after the relevant "bucket" was filled, and the condition in which the feedback was present.

Task	**Feedback**	**Condition**
Transfer	*"It is important to effectively communicate crossings"*	Private and Public
Acknowledge	*"It is important to confirm at appropriate times"*	Private
Identify	*"It is important to identify targets as quickly as possible"*	Private and Public
Assessment	*"Remember to assess the threats posed by all crossing targets"*	Private and Public

All feedback was given using GIFT (Sottilare et al., 2012), as was done in the STT, in response to the comparisons between the player actions and the relevant GIFT domain knowledge file (DKF), wherein individual tasks were evaluated against the three individual-level DKFs in the Private feedback condition and against the one team-level DKF in the Public feedback condition. In both cases, the comparison was of individual action against expected action, thereby fulfilling the individual-level assessment dimension which was discussed above. In some sense, this can be interpreted as the having three individual tutors in the Private condition, one per teammate, and having one overall team tutor in the Public condition.

Independent Variables & Manipulation

Feedback privacy. For each experimental session, feedback privacy served as an independent variable with two between-subjects levels wherein either (1) feedback was shown only to the person to whom it applied (heretofore referred to as “private feedback”) or (2) all feedback was shown to everyone on the team (“public feedback”). Participants received feedback on their performance, provided to them in real-time by an intelligent team tutoring system (ITTS). Imagine this manipulation as a tutor individually conferencing each learner on their performance (private feedback condition), or the tutor telling everyone that an issue with an anonymous team member’s performance has surfaced, regardless of its relevance to the team (public feedback condition).

Role naïveté. In the first trial, all participants were nascent in their roles, only having watched a training video before jumping into the task. In the second and third trials, the participants had experienced their present role once (in Trial 2) or twice (in Trial 3). In the fourth trial, either Spotter 1 or Spotter 2 (randomly assigned per experimental session) switched roles with the Sniper. Thus, the Spotter and the Sniper returned to role naïveté, albeit with observations and the previous training video to guide them. Therefore, role naïveté serves as a second, within-subjects independent variable with three levels: (1) total naïveté (Trial 1), (2) experienced (Trials 2 and 3), and (3) partial naïveté (Trial 4 for those who swapped roles). Previous analysis showed no evidence of learning after Trial 2 (Ouverson, et al., 2018), so Trials 2 and 3 were not separated during interpretation.

Teamwork experience. In a survey given prior to the start of the study, each participant’s teamwork experience frequency was recorded. While this was not a manipulated independent variable, this variable served as a mediating variable with three levels: High, Moderate, and Low frequency.

Teammate familiarity. Participants assigned their own teams by registering for the study as random individuals or with friends. Thus, varying levels of prior teammate familiarity were observed and were used as mediating variables. Just prior to the start of the study, a survey assessing baseline relationships within the teammates was given to each participant. For each teammate, the survey asked, “Have you met teammate X?” and the answers to these were recoded to 0 for “No” and 0.5 for “Yes.” The numbers for each teammate were summed to give the three levels of familiarity displayed in Table 14.

Table 14. Teammate familiarity coding choices.

Categories	Score
No familiarity (has met neither teammate)	0
Partial familiarity (has met one teammate)	0.5
Full familiarity (has met both teammates)	1

Trial. There were four trials per experimental session, thus introducing a time variable. This variable was used to examine how experience with the task influenced the dependent variables included below.

Dependent Variables & Metrics

Dependent variables were derived from scores on quizzes given during the post-experiment survey and data collected by the tutor during the experiment, as shown in Table 15. Each of these is discussed below.

Table 15. Dependent variables examined in this paper.

Dependent Variable	Metric	Frequency
Sniper Goal Awareness	Proportion correct on sniper goal quiz (0.0 – 1.0)	Post-experiment (1x)
Shared Role Awareness	Similarity to teammates on sniper and spotter goal quizzes (0.0 – 1.0)	Post-experiment (1x)
Team Task Awareness	Spearman rank correlation of task quiz answers to correct answers (0.0 – 1.0)	Post-experiment (1x)
Communication	Percentage of prompt acknowledges (0-100%)	Each trial (4x)

After all four trials, a post-experiment survey was given. Shared Role Awareness, Sniper Goal Awareness, and Team Task Awareness were derived from answers to three quizzes (sniper goals, spotter goals, and task) in this post-survey as a measure of shared situational awareness. The inclusion of task and role quizzes (e.g., sniper and spotter goal quizzes) mirrors the efforts of Sætrevik and Eid (2014), who use expert accounts of teamwork to measure shared situational awareness. The quizzes in the present study were based off the tutorial which was given to all participants at the beginning of the study.

Participants were given a list of actions (shown in Table 16), and were asked to identify which actions were goals of the Spotters in the task and the goals of the Sniper in the task. The similarity of each participant's answers to their team mate's answers for each question on both the sniper and spotter goal quizzes makes up the Shared Role Awareness score, while Sniper Goal Awareness is simply the score on the sniper goal quiz.

Table 16. Items from quizzes given after the experiment which were scored compared to teammates' answers to derive Shared Role Awareness and Sniper Goal Awareness.

What are the Goals of the Sniper in this Task?

- ☐ To identify targets new to their zone
- ☐ To keep count of how many targets have left and entered their zone
- ☐ To identify targets leaving their zone
- ☐ To keep count of how many OPFOR are on the map
- ☐ To assess the treats posed by targets
- ☐ To keep count of how many civilians are on the map
- ☐ To acknowledge what their teammates say
- ☐ To count the number of OPFOR wearing vests

What are the Goals of the Spotters in this Task?

- ☐ To assess the treats posed by targets
- ☐ To keep count of how many civilians are on the map
- ☐ To keep count of how many targets have left and entered their zone
- ☐ To count the number of OPFOR wearing vests
- ☐ To keep count of how many OPFOR are on the map

Additionally, participants were given a list of statements of steps to the task for the task quiz (Table 17). They then sorted statements into two categories depending on whether they occurred in the task or not and were asked to order the steps of the task correctly. By finding the Spearman Rank-Order correlation of each participant’s answers, the Team Task Awareness score was derived.

Table 17. The Task Quiz answer key, a list of steps to the task and the order in which they should occur. Participants received these steps in presented in random order and were asked to order them and mark certain ones as erroneous. "NA" flags the erroneous steps in this answer key.

Task steps	Order
Spotter 1 sees a target approaching the 1 pole	1
Spotter 1 transfers a target by pressing the 1 key	2
Spotter 1 transfers a target by pressing the E key	NA
Spotter 2 acknowledges his/her teammate’s communication by pressing the E key	3
Spotter 2 acknowledges his/her teammate’s communication by pressing the 1 key	NA
Spotter 2 sees a target near by the 1 pole	4
Spotter 2 identifies that a target has entered his/her zone by pressing the SPACEBAR key	5
Spotter 2 identifies that a target has entered his/her zone by pressing the E key	NA
Spotter 2 identifies that a target has entered his/her zone by pressing the 1 key	NA
Spotter 2 informs Sniper that a target has entered his/her zone	6

Table 17. (continued)

Task steps	Order
Sniper acknowledges his/her teammate’s communication by pressing the E key	7
Sniper searches for a target in Spotter 2’s zone in the direction of the 1 pole	8
Sniper spots a target and assesses the threat posed by the target	9
Sniper believes target to be a civilian and presses the C key	NA
Sniper believes target to be a civilian and presses the X key	NA
Sniper believes target to be a civilian and presses the Z key	10

Communication was demonstrated through the percentage of prompt acknowledges. This was chosen as the metric because while teams were instructed to use Acknowledgment at certain points in the action sequence, this action was not pivotal to the team’s end goal: assessing the threat posed by potential OPFOR in the environment. Therefore the Acknowledge action was similar to the secondary communication used to develop group affinity (Nardi, 2005) and served as a proxy measure of the communicativeness of the team.

Data Analysis

Hypothesis testing was done by fitting four linear mixed-effects models using the restricted maximum likelihood (REML) criterion which were generated in RStudio using the lme4 package (Bates et al., 2015). Estimated Marginal Means were calculated using the emmeans package (Lenth, 2019). This approach, rather than the standard ANOVA and its variants, was used to account for the fact that individuals in teams cannot be considered independent of one another, so there was additional error for which the researchers must account. Figures showing these differences use a black dot for each of the estimated marginal means, which are centered on their confidence intervals (represented as bars). The difference between the estimated marginal means is not significant if the red arrows in the chart overlap with one another. This overlap is tested for significant difference using a t-test and corresponding p-values.

Tukey's Honest Significant Difference (HSD) was used for multiple comparisons in pairwise differences, and Cohen's *d* values were calculated for each pairwise difference to indicate the size of the effect as a function of the standard deviations of the groups being compared. Cohen (1988) indicated that, when interpreting effect sizes, $0.2 \leq d < 0.5$ showed a small effect, $0.5 \leq d < 0.8$ could be considered a medium-sized effect, and $d \geq 0.8$ showed a large effect.

Results

In analyzing the data, four linear mixed-effects models (LMMs) were created – one for each of the dependent variables, as shown in Table 18.

Table 18. Linear Mixed-effects Models (LMMs) and the hypotheses tested using them

Linear Mixed-effects Model	Hypotheses Tested
Shared Team Awareness = Role naïveté + Random effect(Team)	H1
Sniper Goal Awareness = Role naïveté + Random effect(Team)	H1
Shared Role Awareness = Role naïveté + Team experience frequency + Feedback privacy + Random effect(Team)	H1, H3, H5
Communication = Teammate familiarity + Feedback privacy + Trial + Random effect(Team)	H2, H4, H6

The first two models expressed Shared Team Awareness and Sniper Goal Awareness as a function of role naïveté (total naïveté, experienced, and partial naïveté) and a random effect of the team to which each participant belonged. The third model expressed Shared Role Awareness as a function of role naïveté (total naïveté, experienced, and partial naïveté), the level of self-reported team experience frequency (low, moderate, or high), feedback privacy (Public or Private), and a random effect of the team to which each participant belonged. The last model expressed Communication as a function of teammate familiarity (0, 0.5, 1), feedback privacy (Public or Private), trial (1 – 4), and a random effect of the team to which each participant belonged. The main effects for all models were evaluated by way of the estimated marginal means, calculated using the emmeans package in R Studio (Lenth, 2019).

Does experiencing a new role impact Shared Situational Awareness?

To evaluate Hypothesis 1 – participants who experience more than one role will have higher (A) Team Task Awareness (B) Sniper Goal Awareness and (C) Shared Role Awareness – estimated marginal means were calculated from the LMMs. A purple bar indicates the confidence interval for the estimated marginal mean (black dot) of the measures of Shared SA by each level of the factor, Role Switch, while the arrows can be used to understand statistical significance: if the two arrows overlap with one another, the difference

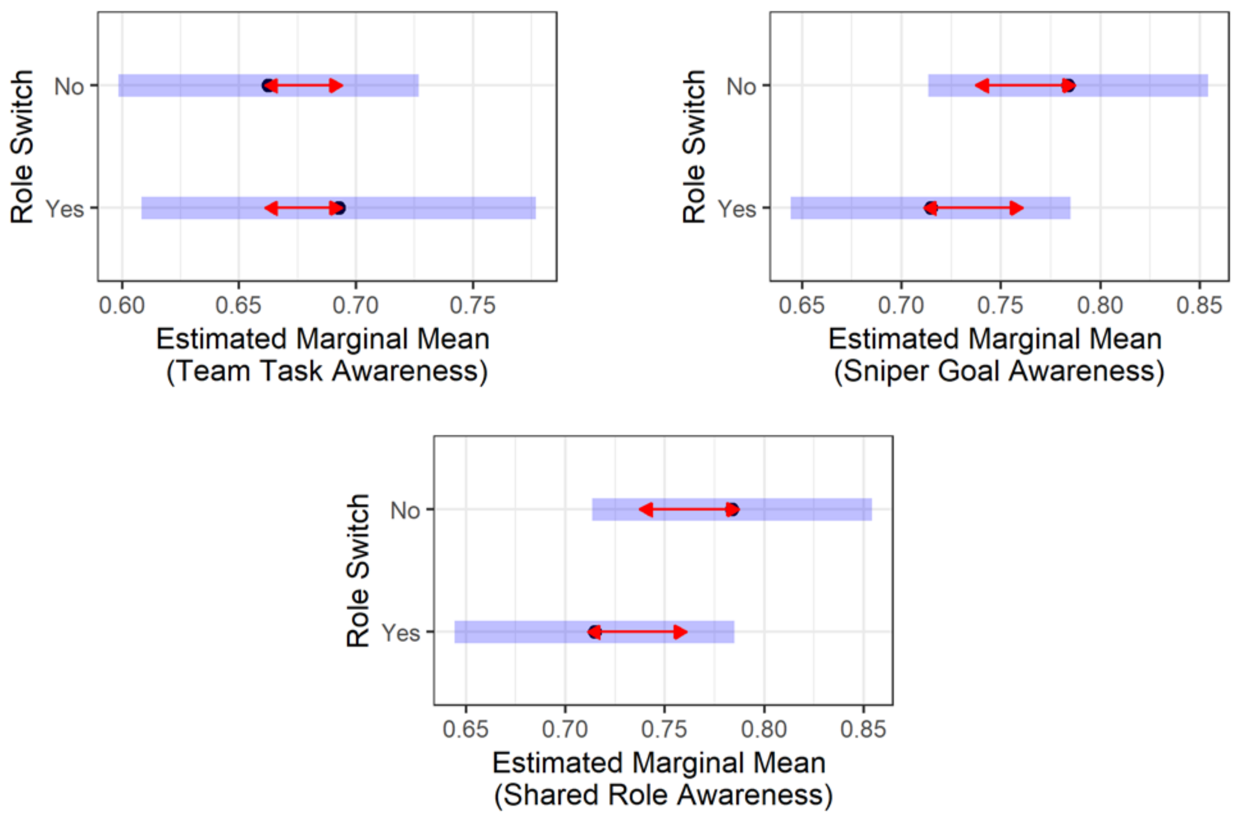

Figure 15. Differences between the estimated marginal means (EMMs) of Shared Situational Awareness by whether the participant switched roles. If the arrows overlap with one another, the difference between the EMMs (black dots centered on the confidence interval bar) is not significant.

between conditions is not statistically significant. No statistically significant difference was found for Team Task Awareness ($t(73) = -0.62$, $p = .54$, $d = 0.12$) or Shared Role Awareness ($t(66) = 0.95$, $p = .35$, $d = -0.11$) between participants who did not switch roles in Trial 4 and those who did. Spotters who did not switch roles were not significantly different on scores of Sniper Goal Awareness ($t(36) = -1.45$, $p = .15$, $d = 0.31$) from those who did switch roles in Trial 4. Results are visualized in Figure 15.

Does Teammate Familiarity impact the ability to communicate?

To evaluate Hypothesis 2 – participants who are familiar with at least one teammate will have fewer Acknowledgment errors – estimated marginal means were calculated from the LMM. No statistically significant difference was found between those with no familiarity and partial teammate familiarity ($t(263) = -1.12$, $p = .50$, $d = -0.14$), partial familiarity and full teammate familiarity ($t(93) = 1.49$, $p = .30$, $d = 0.43$), or full familiarity and no familiarity ($t(71) = 0.90$, $p = .64$, $d = 0.28$).

Does teamwork experience impact Shared Situational Awareness?

To evaluate Hypothesis 3 – persons who work in teams more often score higher on Shared Role Awareness – estimated marginal means were calculated from the LMM. No statistically significant difference was found between High and Low frequency ($t(73) = -0.09$, $p = .99$, $d = 0.05$), High and Moderate frequency ($t(71) = 0.92$, $p = .63$, $d = 0.33$), or Low and Moderate frequency ($t(70) = 1.09$, $p = .52$, $d = 0.29$).

Does role naïveté impact communication?

To evaluate Hypothesis 4 – in Trials 1 and 4, Acknowledge errors for participants will be higher than in Trials 2 and 3 – estimated marginal means were calculated from the LMM. Statistically significant differences in Acknowledge errors were found between Trials 1 and 2 ($t(280) = -4.28$, $p < .001$, $d = -0.61$), Trials 1 and 3 ($t(280) = -4.18$, $p < .001$, *d* = -0.62), Trials 2 and 4 ($t(280) = 4.33$, $p < .001$, $d = 0.58$), and Trials 3 and 4 ($t(280) = 4.23$, $p < .001$, $d = 0.59$), but not between Trials 1 and 4 ($t(280) = 0.12$, $p = 1.00$, $d = 0.06$) or 2 and 3 ($t(280) = 0.09$, $p = 1.00$, $d = 0.01$). Table 19 shows the estimated marginal mean Acknowledgment percentages for each of the trials.

Table 19. Estimated marginal mean (EMM) acknowledge percentages for Trials 1 through 4.

Trial	Acknowledge percentage (EMM)	Confidence Interval	
1	28%	20%	35%
2	43%	35%	51%
3	43%	35%	50%
4	27%	19%	35%

Does Feedback privacy impact Shared Situational Awareness? Does it impact communication?

To evaluate Hypothesis 5 – public feedback will result in higher Shared Role Awareness than private feedback – estimated marginal means were calculated from the LMM. Public feedback did not result in a statistically significant difference for Shared Role Awareness ($t(35) = 1.03$, $p = .31$, $d = -0.29$).

To evaluate Hypothesis 6 (public feedback will result in lower Acknowledgment errors), estimated marginal means were calculated from the LMM. There was no statistically significant difference between public and private feedback for Acknowledgment errors ($t(30) = 1.34$, $p = .19$, $d = -0.28$). The trend of the difference is illustrated in Figure 16.

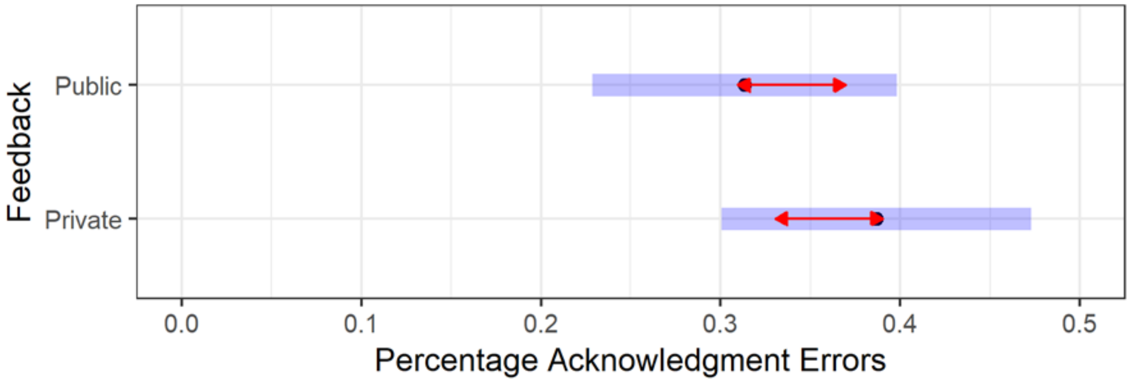

Figure 16. *Differences between the estimated marginal means (EMMs) of the percentage of Acknowledgment Errors by the type of feedback received. If the arrows overlap with one another, the difference between the EMMs (black dots centered on the confidence interval bar) is not significant.*

Discussion

This paper describes the impact of an ITTS on team outcomes of communication and shared SA. While results were generally not aligned with the researchers' predictions, there is a measure of useful information to be gleaned from the described experiment.

Shared Situational Awareness

The first unexpected results are that shared SA does not appear to be influenced by role experience or prior teamwork experience. Previous work has argued that SA is formed from interactions with the environment, including interactions with teammates and the knowledge and abilities identified in oneself (Sætrevik & Eid, 2014). It does appear that shared SA may be somewhat affected by the experience in that role, at least in terms of understanding the goals of the other roles on the team. While spotters who did not switch roles did not score significantly lower on Sniper Goal Awareness, there was evidence of some difference in Sniper Goal Awareness scores as the effect size (d = 0.31) was larger than those for the other two shared SA measures (effect sizes of d = 0.12 for Team Task Awareness and d = -0.11 for Shared Role Awareness) and the result was approaching significance (p = .15). The lack of statistical significance could be due to small sample size or could be a side effect of the high workload introduced by the task, as documented in Ouverson et al. (2018). Other than sample size, there are several potential explanations for this result not aligning with previous research.

One explanation is the measures of shared SA themselves. The answers to the quizzes were drawn from the tutorial given at the beginning of the experimental session. However, participants may have given more weight to the actual actions of the Sniper they interacted with as opposed to the actions described at the outset of the experiment when considering what the goals of that role were. In fact, scores on the measures of shared SA were all quite low, signifying that either the task was hard to understand or the measures were not adequately discriminatory.

Another potential explanation stems from the SwS tutor's disregard of backup behavior when offering feedback for actions. Backup behavior is the taking over of tasks for teammates when they are in need (Burke, Sottilare, Johnston, Sinatra, & Salas, 2017; McIntyre & Salas, 1995), and due to the complexity of attributing them correctly, backup behavior actions were counted as errors in the SwS and resulted in feedback discouraging such actions. In the future, conditionals could be create to mitigate this, counting errors only in both actions in an interdependent task are missed. Additionally, biometric data, such as electro-dermal activity (EDA) could be used to triangulate moments of need and modify the tutor's understanding of learner behavior.

Communication

This study did not offer conclusive evidence that Teammate Familiarity affected the ability of participants to acknowledge their teammates' communication. The hypothesis was that higher familiarity would result in fewer Acknowledgment errors, since this communication task was not integral to overall task advancement or coordination, even though it was important to the team's overall performance because it is a teamwork action. However, the results suggest several conclusions. First, previous research has suggested that one of the reasons why familiarity is beneficial is that it affords the shedding of communication responsibilities, as teammates who are familiar with one another's needs can anticipate rather than communicate (Smith-Jentsch et al., 2015; Tong et al., 2013). This finding might suggest less communication (more Acknowledgment errors).

One anticipated result that did come from this study was the significant impact of role naïveté on communication. Trials in which participants were adjusting to their roles led to lower communication rates, while those in which participants had some experience were related to fewer errors. This is similar to the pattern in task load noted by Ouverson, et al. (2018).

Worth noting is that scores on communication were quite low, which signifies that the task was hard to accomplish. It is also possible that the floor effect has suppressed the results of the study.

Feedback Privacy

Lastly, feedback privacy was not shown to have a significant impact on shared SA or communication. The researchers identify a reason for the occurrence of these results, which were counter to previous literature. For communication, it is possible that the feedback served as a distraction from each participant's human partners. Human-agent teaming has been receiving more attention in the literature, and there is an element of it present in the SwS task. In this task, the tutor must vie for the attention of participants when giving feedback, and as feedback volume increased, the tutor may have been taking too much of the participants' attention. While adjustments were made to mitigate this natural volume increase, a comparison of the average count of feedback in each trial between conditions showed that public condition trials had significantly more feedback (M = 10) than private condition trials (M = 9; $t(255) = 4.31$, $p < .001$). Similarly, this also have affected the shared SA of the participants, especially if the tutor's messages were ignored. The researchers are assuming that the feedback was noticed, as 93% (n = 103) of participants reported having looked at the feedback.

Notably, this experiment does not feature a feedback-free control condition. A control condition was not included in an effort to decrease the number of required participants. Because a control condition was not included in this iteration of testing, the tutor effectiveness (i.e., whether or not the tutor improved performance better than regular practice within the scenario) cannot be accurately evaluated. However, the analysis presented in Ostrander et al. (2019) shows that for a nearly identical scenario and tutor framework combination, the presence of the ITTS resulted in behavioral change of the team members. From this result, it is assumed that the feedback has some effect over just practice in the scenario.

Conclusions and Future Directions

This paper conducts an initial exploration of how team skills are affected by the presence of an ITTS, as well as individual teammate differences in team member familiarity and both face-to-face and virtual team experience. Parallel to preliminary work which examined the workload changes for this experiment (Ouverson et al., 2018), communication improved in Trial 2 and 3 over Trial 1, while Trial 4 resulted in slightly lower performance due to the presence of the role switch. However, results related to shared SA and communication with regards to feedback, past experience, and familiarity were largely inconclusive.

Future work should be conducted to explore a number of questions revealed in this study. First, while the use of keystroke analysis reveals some areas for further development in how the tutor evaluates user performance, more information could come from a comparison to the verbal utterances of the team members. For instance, comparing the performance of teams that use long phrases with those who are straight-to-the-point under various feedback conditions would give more information about the role of familiarity and communication style in teamwork performance. This could be useful in directing team onboarding efforts, and could change the way that team tutoring begins in an ITTS.

Second, better establishing the expectations of the team task by experts or ITTS authors would improve the development of SA evaluation tools, and to fully understand how team shared SA changes over time, this should be measured repeatedly in the experiment, potentially by administering them after each trial. With robust measures of SA and repeated measurement, tutor performance can be better evaluated. Team performance on tasks gives an understanding of technical knowledge gained, but by knowing how a team's understanding of the task and the environment of each teammate changes over time, their performance on teamwork can be evaluated.

Third, with more understanding of the types of feedback which are most impactful for each variety of team task may be needed to shed light in this area. Some work has already been done to understand such a taxonomy (Bonner et al., 2014), and future work should incorporate this information may be used to better adapt feedback to the subtask being evaluated. Finally, additional exploration should attend to whether the feedback is being attended and used, either through the use of a control condition or examination of eye-tracking data, so that feedback effectiveness can be evaluated.

Team training is becoming more virtual to keep pace with the changes in work and provide options for training for rare or dangerous events. As more studies are conducted to evaluate team training which utilizes an ITTS, this work will serve a foundational role in exploring the impact of an ITTS on team skill development. As the second iteration of a tutor developed using a scalable team surveillance task environment (STT to SwS), this work showcases a platform which may be continually improved upon and used to develop and test team training.

CHAPTER 6. ADDITIONAL RESEARCH

The research presented in this thesis raises more questions than just those reported in Chapter 5. These hypotheses are reviewed in Table 20. In addition to questions about outcomes related to teammate interaction, there is information to be gleaned related to the performance of the teams, the performance of the teammates in their roles, and the collective efficacy of each member in Table 20. Descriptive results in the form of histograms will be reported for each of the dependent variables. Next, the models which were tested to establish support for or against the hypotheses are described. Finally, the results of the tests run on those models are detailed and are discussed according to the questions answered by the analyses. These questions are: "Does feedback privacy impact performance?" "Does experience impact performance or collective efficacy?" and "Does familiarity impact performance or collective efficacy?"

Table 20. Hypotheses tested in this section.

	Hypothesis
	Feedback
H7A:	Teams receiving private feedback will perform better than teams receiving public feedback
H7B:	Private feedback will result in higher individual performance than Public feedback
H8:	Participants who use or find the feedback more helpful will have higher individual performance than those who do not use the feedback or do not find it as helpful
	Previous Experience (with teamwork and cooperative-play video games)
H9:	Persons who play a higher proportion co-op video games and play video games more often will have higher overall collective efficacy
H10:	Participants with more frequent video game experience will have higher individual performance than those with less experience
H11:	Participants with more frequent team experience will have a higher overall collective efficacy
H12:	Participants with more frequent team experience will have higher individual performance than those with less experience

Table 20. (continued)

	Hypothesis
	Previous Experience (with teamwork and cooperative-play video games)
H13:	Participants with experience in their role will have higher individual performance than participants with partial naïveté
	Familiarity (with teammates, role and task)
H14A:	Participants who are more familiar with their teammates will have higher individual performance than those who are less familiar
H14B:	Teams with members who are fully familiar with one-another will perform better than teams with no and partial familiarity
H15:	Collective efficacy will increase across Trials 1 through 4

Descriptive Results and Discussion

Before reporting the experimental results, descriptive results for each of the dependent variables were gathered via an examination of the histograms. Interpretation of each graph is provided in addition to the histograms.

Sniper Goal Awareness. Sniper Goal Awareness scores for the sample were not normally distributed, as shown in Figure 17. The data indicate a slight ceiling effect, where 44 participants received perfect scores. A substantial number of participants (*n* = 45) scored

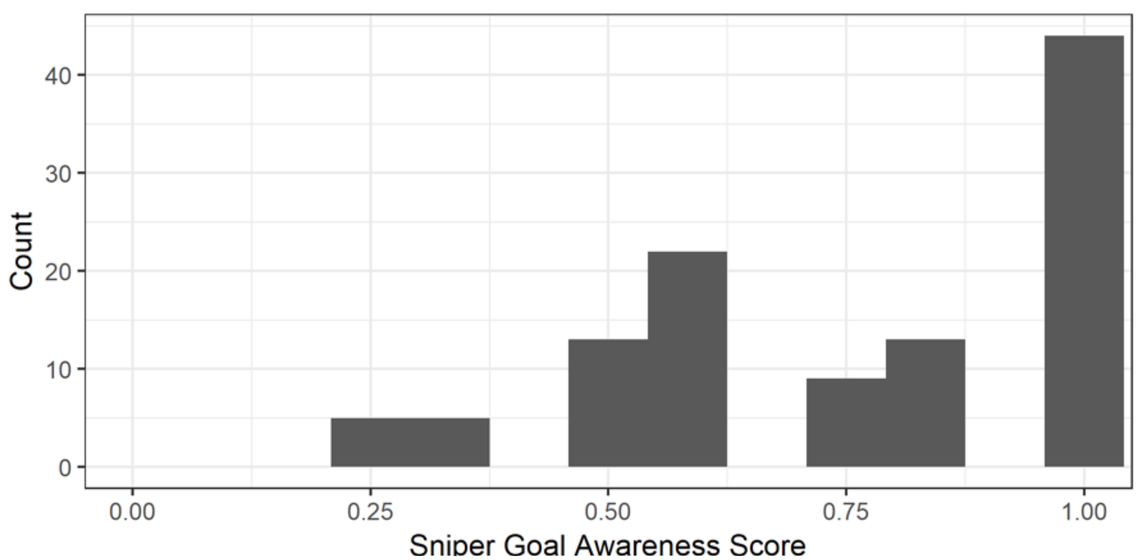

Figure 17. Distribution of Sniper Goal Awareness scores across the participant sample (n = 111).

lower than 60%. It was expected that primary role Snipers (*n* = 37) would score high, and primary role Spotters who never switched (*n* = 37) would score low.

Shared Role Awareness. Shared Role Awareness scores for the sample were less skewed than the Sniper Goal Awareness scores, but still show a left skew. These data more closely resemble a normal distribution, as shown in Figure 18. There is a slight ceiling effect, with the majority of participants scoring above 70%. It was expected that public feedback would increase the similarity between teammate understanding of each team member's role, so just over half of the participants (*n* = 57, those who received public feedback) should score higher than their private-feedback-receiving counterparts.

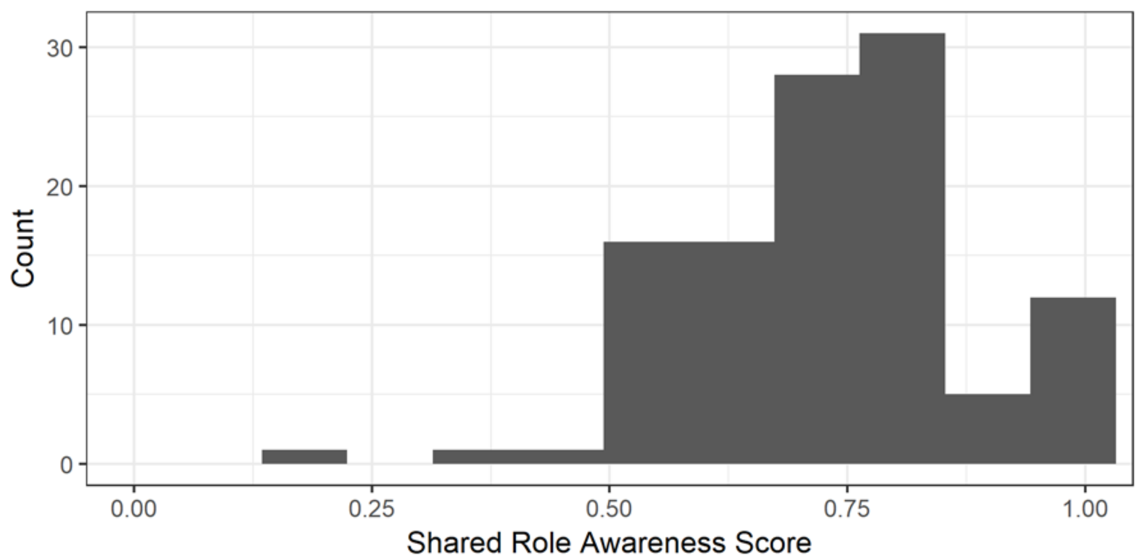

Figure 18. Distribution of Shared Role Awareness scores across the participant sample (n = 111).

Team Task Awareness. Team Task Awareness scores for the sample are again skewed to the left, as shown in Figure 19. This skewing may suggest that the Team Task Awareness quiz was too easy, since most participants scored well.

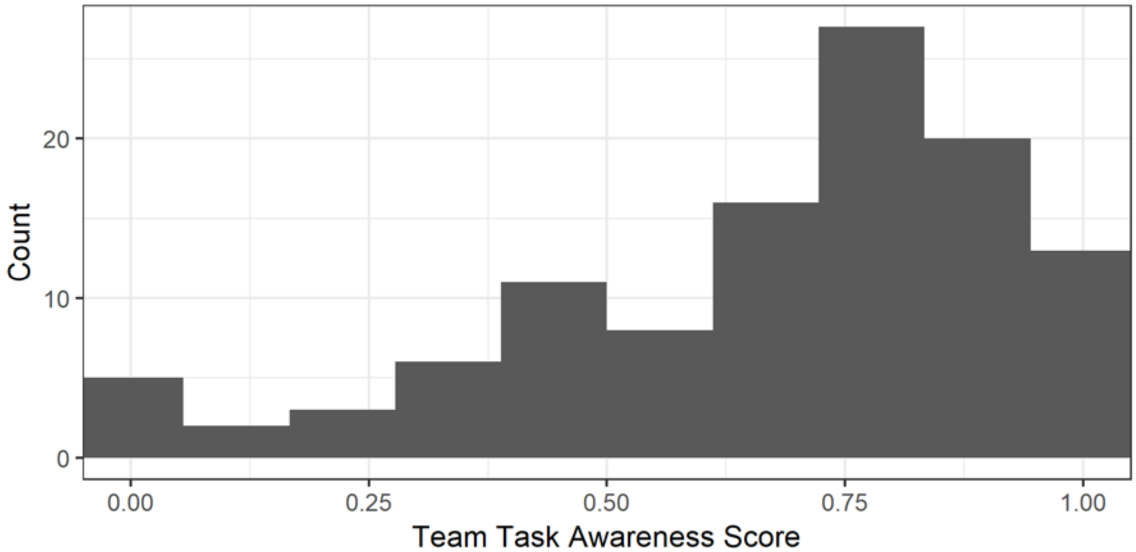

Figure 19. Distribution of Team Task Awareness scores across the participant sample (n = 111).

Collective efficacy. Collective efficacy scores were fairly normally distributed, as shown in Figure 20. This indicates that the sample is representative of the population.

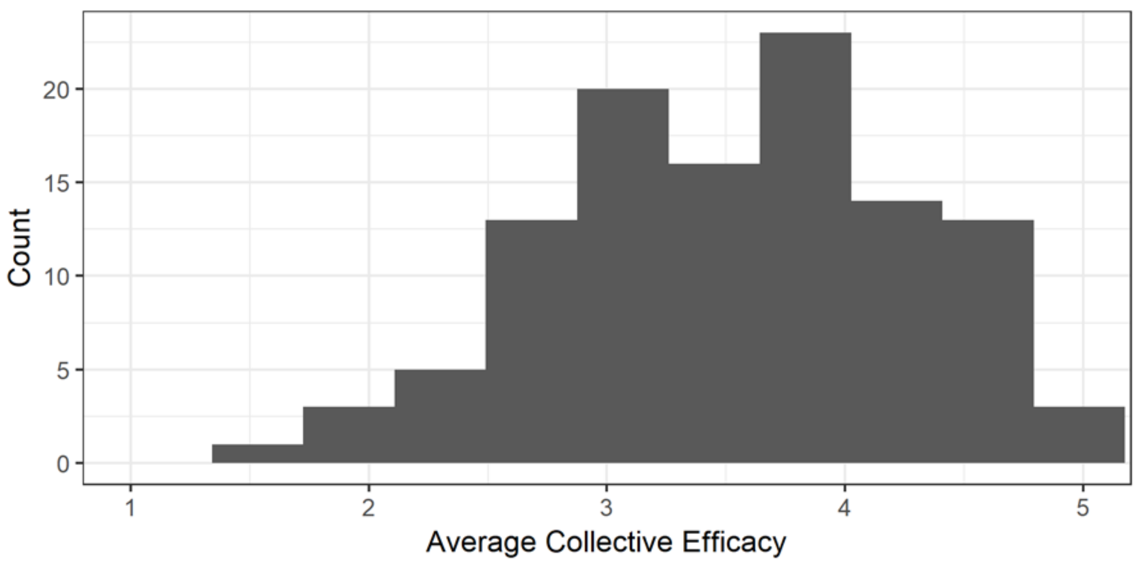

Figure 20. Distribution of collective efficacy scores averaged over trial across the participant sample (n = 111).

Team performance. Team performance was highly right skewed, indicating a floor-effect in the data. The majority of participants scored less than 20% during the trials, as shown in Figure 21. These data indicated the overall coordination of the task was highly difficult, but it may instead indicate an issue in the method of scoring or a lack of engagement from the participants.

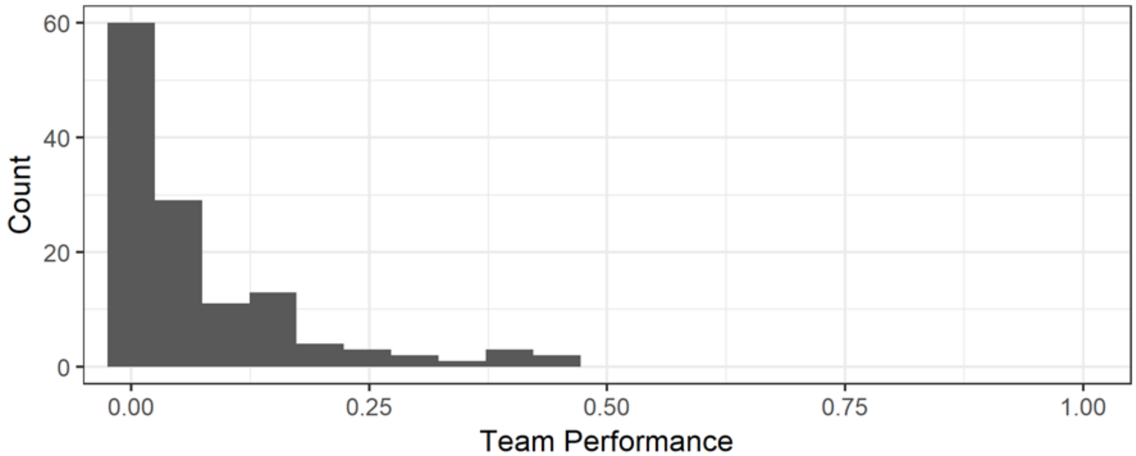

Figure 21. Distribution of team performance scores across the team sample (n = 37) and over all four trials.

Individual performance. Individual performance is a count of the missed Transfers, Identifies, Acknowledges, and Assessments of each participant. In this way, individual performance is not a metric that can be represented as a single number. The distributions of these scores across the sample, summing across trials, are shown in Figure 23. The general trend for all of these distributions is a rightward skew, which aligns with the observed distribution of the team performance scores. By examining these histograms, one concludes that it is likely that the Identify action, followed by the Assessment action are most negatively affecting the overall team performance score. Furthermore, the bimodality of Identifies could be due to the lack of feedback given for identifies in the public condition, which was done in an attempt to reduce the overall feedback given when feedback audience moves from individual to team level.

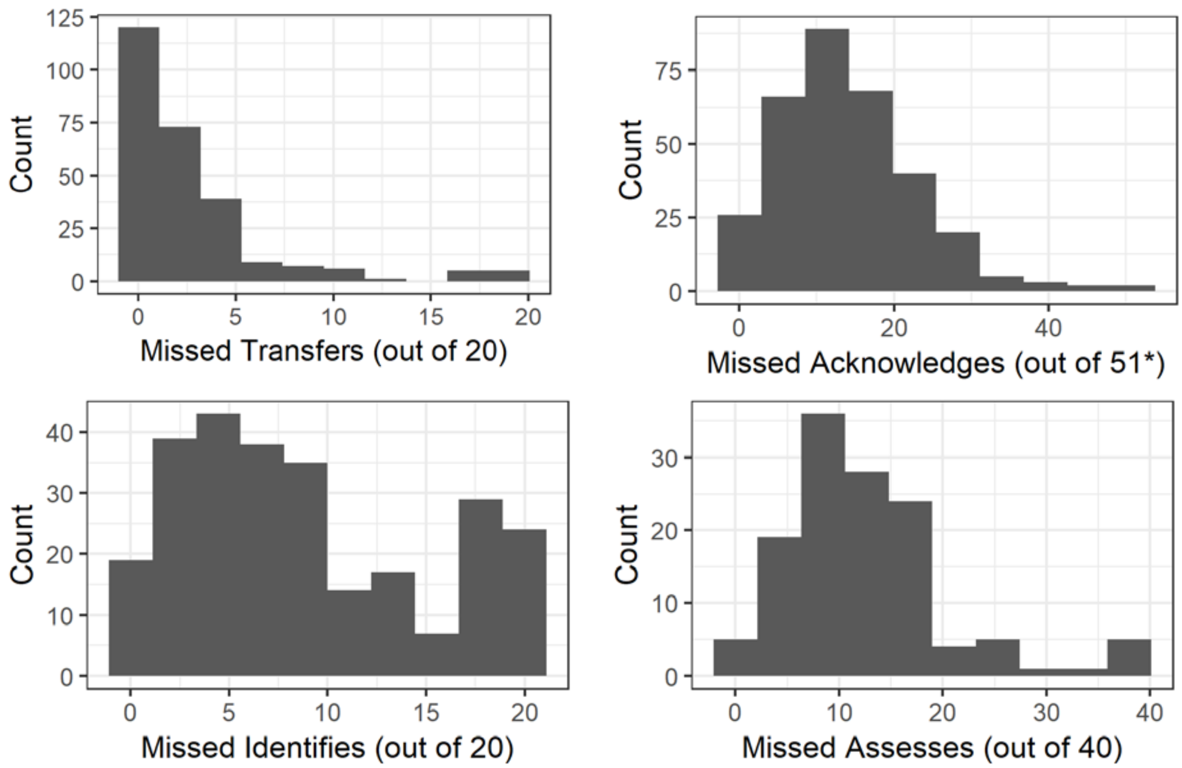

*Figure 23. Distributions of the four measure of individual performance across the participant sample (n = 111) and over the four experimental trials. *dependent on the player Transfer and Identify actions, not on simulation events*

Communication. Communication is operationalized as the percent of correct Acknowledge actions. This was chosen as the measure of communication because the Acknowledge action is used to inform a teammate that they have been heard, and even

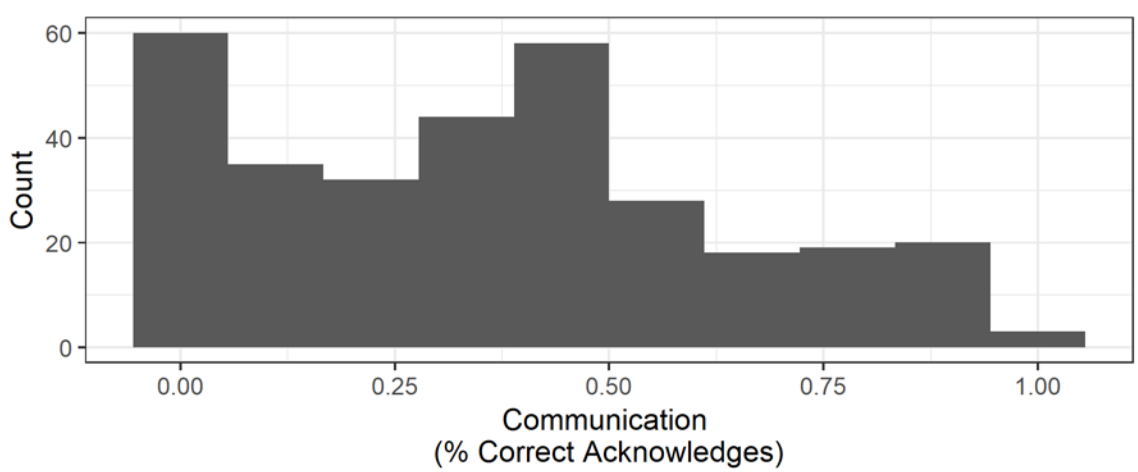

Figure 22. Distribution of communication scores across the participant sample (n = 111) over the four trials.

without these, the task would continue, more or less. Additionally, Acknowledge is examined as a percentage and a count do to the reliance of this action on the Transfer and Identify actions of the player's teammates. The distribution of communication scores is highlighted in Figure 22.

Models

To answer questions related to these variables, linear mixed effects models were created using the lme4 package in R Studio (Bates et al., 2015). The models shown in Table 21 were created to evaluate different impacts on individual performance. For each of these actions except acknowledges, only two of the three participants have the chance to make errors, and each individual's actions are kept separate. When comparisons are made between roles, the errors compared are an average of those over Trials 1-3 by the primary role holder and those made in Trial 4 by the secondary role holder.

Table 21. Linear Mixed-effects Models (LMMs) and the hypotheses tested using them. All response variables are collected at the individual level except Team Performance.

Linear Mixed-effects Model	Hypotheses Tested
Missed Transfers = Feedback privacy + Primary role + Video game-play frequency + Team experience frequency + Feedback use + Random effect(Team)	H7B, H8, H10, H12, H13, H14A
Missed Acknowledges = Feedback privacy + Primary role + Video game-play frequency + Team experience frequency + Feedback use + Random effect(Team)	H7B, H8, H10, H12, H13, H14A
Missed Identifies = Feedback privacy + Primary role + Video game-play frequency + Team experience frequency + Feedback use + Random effect(Team)	H7B, H8, H10, H12, H13, H14A
Missed Assessments = Feedback privacy + Primary role + Video game-play frequency + Team experience frequency + Feedback use + Random effect(Team)	H7B, H8, H10, H12, H13, H14A
Team Performance = Feedback privacy + Teammate familiarity + Random effect(Team)	H7A, H7, H14B
Collective Efficacy = Trial + (Video game-play frequency)*(Cooperative video game-play) + Random effect(Team)	H9, H11, H15

The first four models expressed Missed Transfers, Missed Acknowledges, Missed Identifies, and Missed Assessments as a function of feedback privacy (Public or Private), primary role (Spotter or Sniper), the level of frequency of self-reported video game play (low, moderate, or high), the level of self-reported team experience frequency (low, moderate, or high), the familiarity of each participant with their team members (none, partial, or total), the self-reported use and appreciation of the feedback (ignored, distracting, not helpful, somewhat helpful, or helpful), and a random effect of the team to which each participant belonged.

The fifth model expressed Team Performance as a function of feedback privacy (Public or Private), teammate familiarity (0, 0.5, 1), and a random effect of the team to which each participant belonged. The last model expressed Collective Efficacy as a function of trial (1 – 4) and an interaction between the level of frequency of self-reported video game play (low, moderate, or high) and of cooperative video game play (none, low, or high), and a random effect of the team to which each participant belonged. The main effects for all models and interaction effect for the last model were evaluated by way of the estimated marginal means, calculated using the emmeans package in R Studio (Lenth, 2019).

Results

Does feedback privacy impact performance?

To evaluate whether teams (H7A) and participants (H7B) receiving private feedback performed better than those receiving public feedback, the estimated marginal means were evaluated for the feedback factor. No statistical difference was found between Public (*EMM* = 0.08, *CI* = 0.03, 0.13) and Private feedback (*EMM* = 0.07, *CI* = 0.01, 0.12) for team performance ($t(28) = -0.52$, $p = .61$, $d = -0.17$), thus hypothesis H7A was rejected. For individual performance, only Acknowledge errors saw a statistical difference between Private (*EMM* = 13.8, *CI* = 10.4, 17.2) and Public (*EMM* = 18.0, *CI* = 14.7, 21.2) feedback, as shown in Table 22; however, the size of the effect was negligible. Therefore, Hypothesis

7B was partially supported by the data; the hypothesis cannot be considered fully rejected or supported.

Table 22. Effect of feedback condition on individual performance, as measured by errors.

	Errors			
Contrast	**Transfer**	**Acknowledge**	**Identify**	**Assess**
Private - Public	$t(28) = -1.25$, $p = .22$, $d = -0.33$	$t(28) = -2.28$, $p = .03$*, $d = -0.16$	$t(28) = 0.38$, $p = .71$, $d = -0.15$	$t(23) = 0.39$, $p = .70$, $d = 0.27$
* significant at α = .05				

To evaluate whether participants had higher individual performance if they found the feedback more helpful (H8), the estimated marginal means were examined for each factor and each error source. As shown in Table 23, this hypothesis was partially supported by the data. Transfer errors were statistically significantly higher for individuals who did not use the feedback (*EMM* = 2.93, *CI* = 1.18, 4.69) as compared to individuals who used it, but found it distracting (*EMM* = 0.32, *CI* = -1.68, 2.32).

Table 23. Hypothesis tests for the effect of feedback use on individual performance. Multiple-comparison adjustments done using Tukey HSD.

	Errors			
Contrast	**Transfer**	**Acknowledge**	**Identify**	**Assess**
Ignored-Distracting	$t(164) = 3.03$, **$p = .02$*,** $d = 0.34$	$t(185) = -0.34$, $p = 1.00$, $d = -0.46$	$t(164) = -0.79$, $p = .93$, $d = 0.25$	$t(50) = 1.43$, $p = .61$, $d = 0.06$
Ignored-Not Helpful	$t(152) = 1.79$, $p = .38$, $d = 0.57$	$t(130) = -2.35$, $p = .14$, $d = -0.48$	$t(138) = 2.38$, $p = .13$, $d = 0.64$	$t(31) = -0.41$, $p = .99$, $d = 0.23$
Ignored-Somewhat Helpful	$t(160) = 1.16$, $p = .78$, $d = 0.14$	$t(137) = -0.55$, $p = 1.00$, $d = -0.71$	$t(154) = -0.11$, $p = 1.00$, $d = -0.24$	$t(41) = 0.63$, $p = .97$, $d = 0.40$
Ignored-Helpful	$t(159) = -0.22$, $p = 1.00$, $d = -0.17$	$t(157) = -0.97$, $p = .87$, $d = -0.04$	$t(151) = 1.99$, $p = .27$, $d = 0.48$	$t(63) = 1.81$, $p = .38$, $d = 0.42$
* significant at α = .05 ¤ significant at α = .10				

Table 23. (continued)

	Errors			
Contrast	**Transfer**	**Acknowledge**	**Identify**	**Assess**
Distracting-Not Helpful	$t(163) = -0.88$, $p = .90$, $d = 0.34$	$t(124) = -1.84$, $p = .35$, $d = -0.06$	$t(135) = 2.72$, $\boldsymbol{p = .06}$¤, $d = 0.32$	$t(36) = -1.86$, $p = .86$, $d = 0.15$
Distracting-Somewhat Helpful	$t(163) = -1.00$, $p = .85$, $d = -0.27$	$t(180) = -0.30$, $p = 1.00$, $d = -0.29$	$t(164) = 0.46$, $p = .99$, $d = -0.45$	$t(58) = -0.47$, $p = .99$, $d = 0.24$
Distracting-Helpful	$t(162) = -2.81$, $\boldsymbol{p = .07}$¤, $d = -0.48$	$t(133) = -0.52$, $p = .98$, $d = 0.43$	$t(154) = 2.29$, $p = .15$, $d = 0.18$	$t(39) = 0.72$, $p = .95$, $d = 0.29$
Not Helpful - Somewhat Helpful	$t(160) = -0.19$, $p = 1.00$, $d = -0.67$	$t(176) = 1.25$, $p = .73$, $d = -0.20$	$t(154) = -1.81$, $p = .37$, $d = -0.96$	$t(53) = 1.05$, $p = .83$, $d = 0.09$
Not Helpful -Helpful	$t(146) = -1.77$, $p = .40$, $d = -0.68$	$t(121) = 1.24$, $p = .73$, $d = 0.45$	$t(130) = -0.43$, $p = .99$, $d = -0.16$	$t(39) = 2.05$, $p = .26$, $d = 0.15$
Somewhat Helpful-Helpful	$t(164) = -1.35$, $p = .66$, $d = -0.29$	$t(145) = -0.17$, $p = 1.00$, $d = 0.68$	$t(163) = 1.54$, $p = .54$, $d = 0.76$	$t(31) = 0.97$, $p = .87$, $d = 0.07$
* significant at $\alpha = .05$ ¤ significant at $\alpha = .10$				

In addition to the difference within Transfer and Identify, there are trends shown in the differences for other combinations and other error sources in Figure 24. For example, there tend to be more errors for participants who reported finding the feedback somewhat helpful, as compared to finding it distracting. A similar statement could be made about the general trend for reporting the feedback as “not helpful.”

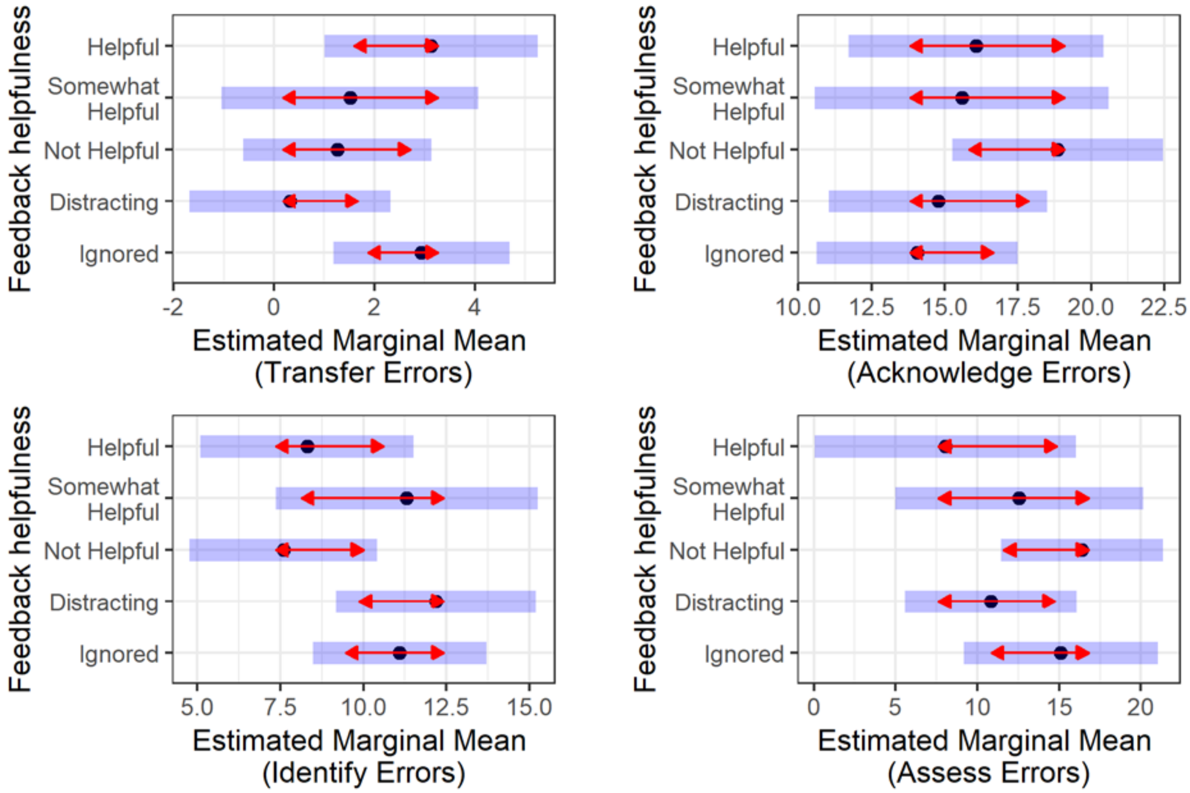

Figure 24. Estimated Marginal Means and confidence intervals for the four measures of individual performance by feedback helpfulness.

Does experience impact performance or collective efficacy?

To evaluate whether participants who play a higher proportion of cooperative video games and play video games more often had higher overall collective efficacy (H9), estimated marginal means for the interactions were calculated and tested. This hypothesis was not supported, but an interesting interaction was uncovered. As shown in Table 24, statistically significant differences were found in collective efficacy between high (*EMM* = 2.18, *CI* = 1.48, 2.88) and low (*EMM* = 3.86, *CI* = 3.54, 4.18) levels of cooperative gameplay experience within low levels of overall video game play frequency. A similar difference was found between high and none (*EMM* = 4.33, *CI* = 3.78, 4.87) levels of cooperative gameplay experience within low levels of overall video game play frequency.

Table 24. Hypothesis tests of the effects of the proportion of cooperative video game play by the level of overall video game play on collective efficacy. Multiple-comparison adjustments done using Tukey HSD.

Contrast	Hypothesis test	Effect size
Video games: High		
High-Low	$t(237) = 0.915$, $p = .63$	$d = 0.18$
Video games: Moderate		
High-Low	$t(259) = 1.23$, $p = .44$	$d = 0.26$
High-None	$t(265) = 0.06$, $p = 1.00$	$d = 1.56$
Low-None	$t(264) = -0.35$, $p = .94$	$d = 1.13$
Video games: Low		
High-Low	$t(170) = -4.40$, $p = .001$**	$d = -1.81$
High-None	$t(154) = -4.67$, $p < .001$**	$d = -0.96$
Low-None	$t(218) = -1.56$, $p = .27$	$d = 0.13$
** significant at $\alpha = .01$ * significant at $\alpha = .05$		

To evaluate whether participants with more frequent video game experience showed higher individual performance than those with less experience (H10), estimated marginal means were calculated and tested for each error source. As shown in Table 25, statistically significant differences in missed transfers were shown between high (*EMM* = -0.52, *CI* = -2.66, 1.62) and low (*EMM* = 4.43, *CI* = 2.53, 6.33), high and moderate (*EMM* = 1.58, *CI* = -0.08, 3.25), and low and moderate levels of frequency of previous video game experience. Higher frequencies of video game play were shown to result in lower Transfer errors, on average. A marginally significant effect was found for Assessment errors between high (*EMM* = 9.03, *CI* = 2.68, 15.4) and low (*EMM* = 17.19, *CI* = 11.38, 23.0) frequencies of previous video gameplay experience, showing a similar trend for the sniper role. Therefore, Hypothesis 10 was partially supported by the data; the hypothesis cannot be considered fully rejected or supported.

Table 25. Hypothesis tests of the effects of video game play frequency level on errors. Multiple-comparison adjustments done using Tukey HSD.

	Errors			
Contrast	**Transfer**	**Acknowledge**	**Identify**	**Assess**
High-Low	$t(147) = -4.72$, $p < .001$**, $d = -0.02$	$t(148) = -1.62$, $p = .24$, $d = -0.34$	$t(139) = 0.44$, $p = .90$, $d = 0.20$	$t(64) = -2.31$, $p = .06$¤, $d = -0.09$
High-Moderate	$t(120) = -2.45$, $p = .04$*, $d = -0.43$	$t(106) = -1.01$, $p = .57$, $d = 0.25$	$t(105) = 1.01$, $p = .57$, $d = 0.21$	$t(52) = -0.79$, $p = .71$, $d = -0.30$
Low-Moderate	$t(164) = 3.08$, $p = .01$**, $d = -0.41$	$t(153) = 0.96$, $p = .61$, $d = 0.11$	$t(161) = 0.42$, $p = .91$, $d = 0.02$	$t(42) = 1.94$, $p = .14$, $d = -0.19$
** significant at $\alpha = .01$ * significant at $\alpha = .05$ ¤ significant at $\alpha = .10$				

To evaluate whether participants with more frequent team experience had a higher overall collective efficacy (H11), estimated marginal means were calculated and tested for each contrast pair. As shown in Table 26, there was a statistically significant difference in collective efficacy between participants who had high (*EMM* = 3.72, *CI* = 3.41, 4.03) or moderate (*EMM* = 3.41, *CI* = 3.11, 3.70) frequencies of team experience and those with low experience (*EMM* = 3.80, *CI* = 3.54, 4.05). Therefore, Hypothesis 11 was partially supported by the data; the hypothesis cannot be considered fully rejected or supported.

Table 26. Hypothesis tests for the effect of the frequency of previous team experience on collective efficacy. Multiple-comparison adjustments done using Tukey HSD.

Contrast	**Hypothesis tests**	**Effect size**
High-Low	$t(244) = 3.34$, $p = .003$**	$d = 0.23$
High-Moderate	$t(262) = -0.04$, $p = 1.00$	$d = 0.11$
Low-Moderate	$t(241) = -4.57$, $p < .001$**	$d = -0.11$
** significant at $\alpha = .01$		

To evaluate whether participants with more frequent team experience will have higher individual performance than those with less experience (H12), estimated marginal means were calculated and tested for each error source. As shown in Table 27, this hypothesis was partially supported by the data. Individuals with a high frequency of previous team experience (*EMM* = 0.05, *CI* = -1.94, 2.1) outperformed those with moderate experience (*EMM* = 3.90, *CI* = 2.29, 5.51) on Transfers, having fewer Transfer errors on average. However, a trend counter to this hypothesis was also uncovered, where moderate frequencies of team experience had more errors that low-frequency individuals (*EMM* = 1.54, *CI* = -0.62, 3.71).

Table 27. Hypothesis test results for the effects of the level of frequency of team experience on errors. Multiple-comparison adjustments done using Tukey HSD.

	Errors			
Contrast	**Transfer**	**Acknowledge**	**Identify**	**Assess**
High-Low	$t(130) = -1.29$, $p = .40$, $d = -0.19$	$t(126) = 0.20$, $p = .98$, $d = 0.15$	$t(113) = 0.73$, $p = .75$, $d = -0.36$	$t(28) = 0.08$, $p = 1.00$, $d = 0.63$
High-Moderate	$t(159) = -4.47$, $p < .001$**, $d = 0.02$	$t(140) = 0.25$, $p = .97$, $d = 0.17$	$t(149) = 0.28$, $p = .96$, $d = -0.22$	$t(29) = 0.50$, $p = .87$, $d = 0.18$
Low-Moderate	$t(120) = -2.60$, $p = .03$*, $d = 0.19$	$t(134) = -0.01$, $p = 1.00$, $d = 0.06$	$t(110) = -0.66$, $p = .79$, $d = 0.12$	$t(46) = 0.46$, $p = .89$, $d = -0.46$
** significant at $\alpha = .01$ * significant at $\alpha = .05$				

Does familiarity impact performance or collective efficacy?

To evaluate whether participants with experience in their role (H13) and whether participants who were more familiar with their teammates (H14A) had higher individual performance than those who had less experience or were less familiar with their teammates, respectively, estimated marginal means were calculated and tested for each error source. As shown in Table 28, Hypothesis 13 was partially supported by the data. Spotters (*EMM* = 13.3, *CI* = 10.7, 15.9) had statistically significantly fewer Acknowledge errors than snipers (*EMM* = 18.4, *CI* = 14.9, 22.0). The data did not support Hypothesis 14A based off the results of the t-test of the estimated marginal means, but Hypothesis 14A was considered partially supported by the data, however, for two reasons. First, the effect sizes for differences between None and Total Familiarity was moderate. Second, the effect sizes for differences between Partial and Total Familiarity was large. This indicates that the tests did not have enough statistical power to uncover the differences via t-test.

Table 28. Hypotheses tests results for the effects of primary role and teammate familiarity on errors. Multiple-comparison adjustments done using Tukey HSD.

	Errors			
Contrast	**Transfer**	**Acknowledge**	**Identify**	**Assess**
Primary Role				
Sniper-Spotter	$t(157) = -1.10$, $p = .27$, $d = 0.23$	$t(201) = 3.26$, $p = .001$**, $d = 0.67$	$t(160) = -0.34$, $p = .74$, $d = -0.04$	$t(65) = -0.72$, $p = .47$, $d = 0.14$
Teammate Familiarity				
None-Partial	$t(130) = -1.98$, $p = .12$, $d = -0.14$	$t(96) = 1.31$, $p = .40$, $d = 0.08$	$t(112) = 1.91$, $p = .14$, $d = 0.18$	$t(37) = -0.38$, $p = .92$, $d = 0.31$
None-Total	$t(32) = 0.25$, $p = .97$, $d = 0.09$	$t(37) = -0.89$, $p = .65$, $d = -0.28$	$t(33) = -1.01$, $p = .58$, $d = -0.57$	$t(19) = -0.07$, $p = 1.00$, $d = 0.31$
Partial-Total	$t(36) = 1.17$, $p = .48$, $d = 0.22$	$t(39) = -1.54$, $p = .29$, $d = -0.41$	$t(36) = -1.93$, $p = .15$, $d = -0.81$	$t(22) = 0.16$, $p = .99$, $d = -0.44$
** significant at $\alpha = .01$				

To evaluate whether teams with members who are fully familiar with one-another performed better than teams with no and partial familiarity (H14B), estimated marginal means were calculated and tested for each contrast. As shown in Table 29, there was no statistical evidence of any difference in team-level performance depending on how familiar the team's members were with one another. Hypothesis 14B was rejected.

Table 29. Hypothesis test results for the effect of team member familiarity level on team performance. Multiple-comparison adjustments done using Tukey HSD.

Contrast	**Hypothesis Test**	**Effect size**
None-Partial	$t(28) = 0.05$, $p = 1.00$	$d = 0.23$
None-Total	$t(28) = -0.44$, $p = .90$	$d = -0.24$
Partial-Total	$t(28) = -0.46$, $p = .89$	$d = -0.26$

To evaluate whether collective efficacy increased across Trials 1 through 4 (H15), estimated marginal means were calculated and tested for each of the contrast pairs. As shown in Table 30, in Trials 2 (*EMM* = 3.70, *CI* = 3.44, 3.97), 3 (*EMM* = 3.88, *CI* = 3.61, 4.14), and 4 (*EMM* = 3.80, *CI* = 3.54, 4.07) participants reported statistically significantly higher collective efficacy than in Trial 1 (*EMM* = 3.19, *CI* = 2.92, 3.45). However, the incremental increase from Trial 2 to Trial 3 was not significant, so the hypothesis is considered supported, although it is only partially supported.

Table 30. Hypothesis test results for the effect of trial on collective efficacy. Multiple-comparison adjustments done using Tukey HSD.

Contrast	**Hypothesis tests**	**Effect size**
1-2	$t(233) = -5.15$, $p < .001$**	$d = -0.56$
1-3	$t(233) = -6.92$, $p < .001$**	$d = -0.76$
1-4	$t(233) = -6.06$, $p < .001$**	$d = -0.69$
2-3	$t(233) = -1.77$, $p = .29$	$d = -0.23$
2-4	$t(233) = -0.94$, $p = .79$	$d = -0.15$
3-4	$t(23) = 0.83$, $p = .84$	$d = 0.08$
** significant at $\alpha = .01$		

Discussion and Conclusions

Several hypotheses regarding the effects of feedback audience, experience with various expected influential activities, and familiarity with team roles and members on individual and team performance as well as collective efficacy were tested. While not all hypotheses were fully supported (see Table 31), the results are informative, nonetheless.

Table 31. Hypotheses tested in this section and the result of the hypothesis tests

	Hypothesis	Supported?
H7A:	Teams receiving private feedback will perform better than teams receiving public feedback	*Partial Support*
H7B:	Private feedback will result in higher individual performance than Public feedback	*Partial Support*
H8:	Participants who use or find the feedback more helpful will have higher individual performance than those who do not use the feedback or do not find it as helpful	*Partial Support*
H9:	Persons who play a higher proportion co-op video games and play video games more often will have higher overall collective efficacy	Rejected
H10:	Participants with more frequent video game experience will have higher individual performance than those with less experience	Rejected
H11:	Participants with more frequent team experience will have a higher overall collective efficacy	*Partial Support*
H12:	Participants with more frequent team experience will have higher individual performance than those with less experience	*Partial Support*
H13:	Participants with experience in their role will have higher individual performance than participants with partial naïveté	*Partial Support*
H14A:	Participants who are more familiar with their teammates will have higher individual performance than those who are less familiar	*Partial Support*
H14B:	Teams with members who are fully familiar with one-another will perform better than teams with no and partial familiarity	Rejected
H15:	Collective efficacy will increase across Trials 1 through 4	**Full Support**

Feedback

First, feedback audience did affect individual performance – specifically on the Acknowledgment errors of participants, although the size of the effect was negligible. Private feedback resulted in fewer missed Acknowledges than did public feedback, and this pattern existed for Transfers, but with a small effect size and no statistical significance. No effect was seen for Identify or Assess errors nor for team performance. While there was some statistical significance, the differences between Public and Private feedback did not garner a large enough effect size for one to conclude that the differences are important to tutor design.

This could indicate that the feedback needs to be better calibrated for the communication and coordination actions measured in this study or that team-level performance may not be so easily measured by a single number meant to represent coordinated action.

Since Identify actions only received private feedback, the absence of effect for that specific action may indicate, more specifically, that the feedback should be redesigned and re-evaluated. Or, it may indicate that the Identify action is representative of a skill which is not best trained with just-in-time feedback. Research is needed to classify types of actions integral to teamwork and the types of feedback that are most effective for each classification.

Self-identified feedback use was shown to impact performance, as those participants who used the feedback and found it more useful saw decreases in errors over participants who did not use the feedback or did not find it as useful. A weak to moderate effect was seen for Transfer and Identify errors, wherein errors were higher for participants who did not use the feedback compared to those who used the feedback selectively, either reporting ignoring it or noting it as distracting. Assess errors were not significantly affected by level of feedback attention, although there was a difference with a weak effect size between participants who ignored the feedback and those who reported it as at least somewhat helpful. The lack of significance of self-reported feedback use could be due to a lack of power due to sample size.

Counter to expectations, Acknowledge errors trended higher, on average, for individuals who paid more attention or found the feedback more helpful. Again, the effects were weak for the difference between not using it and ignoring it or finding it distracting, but the effect size for the difference between not using it and finding it somewhat helpful was large. This could be evidence of an effect of feedback on communication, as the Acknowledge action is fundamentally different from the other actions which are trained by the ITTS. Acknowledgment, at its base, is a communicative action, whereas the other actions are largely task and coordination based. Since feedback is in-itself a form of communication between users and the tutor, it could detract from communication within the teams. More work is needed to evaluate the impact of real-time feedback on task performance, especially in the same modality.

Overall, the relationships between self-reported feedback use and performance are likely an indication of the actual impact of feedback on performance, but more research is needed to elucidate the relationship between users' self-reported attention and eye-tracking data showing attention to feedback within each trial. Additionally, while ignorance of the feedback could be an indication of non-compliance, these data were not excluded because there are a multitude of reasons why someone may not look at the feedback and not many of them are malicious. Supportive of this is the fact that no participants ignored the feedback during all four trials. Lastly, feedback use and communication with human teammates may be at odds with each other, and a future demarcation of the types of tasks and their most effective feedback style would help to solidify this conjecture into theory.

Video game and team experience

Second, experience was shown to affect collective efficacy and individual performance. While not in alignment with the original hypothesis, cooperative video game experience had an interesting effect on collective efficacy. Only participants who played few video games overall showed differences in collective efficacy, depending on how many of those video games involved teamwork. These differences were statistically significant and had large effect sizes. It is expected that this result indicates an individual's teamwork ability is impacted by their video game task skill, as a person who does not play video games very often could be expected to have less skill in the roles of that game, which would then be amplified by the team's success rate.

Video game experience additionally was shown to affect individual experience. Across all error types, except Identifies, higher levels of video game experience were tied to lower average errors. The effect sizes of the differences between high and low frequency were small for Acknowledge errors; the differences between high and moderate frequency were also small for all actions. Identifies showed a break from this general pattern, as high video game experience resulted in higher average errors, although this effect was weak or negligible and not statistically significant. It is possible that playing video games only prepares participants for certain kinds of tasks, although more work to classify the tasks in the SwS and in popular video game genres must be done to be able to test this.

Relatedly, experience with teams affected individual performance and collective efficacy. For individuals who work in teams at a low frequency, collective efficacy is statistically significantly lower than those who work in teams at a moderate or high frequency, although these levels were not significantly different from one another. Additionally, the size of the effect of the difference between high and low frequency was small, while the other two differences were negligible.

As noted in the literature, collective efficacy depends more on the team outcomes than on the frequency of experience in teams (Katz-Navon & Erez, 2005; Tasa et al., 2007). If a person works in teams daily, but these teams make no progress toward accomplishing goals, that person's collective efficacy would be understandably low. However, daily teamwork would hopefully improve performance over time if the team was committed to accomplishing the task. The data were interpreted to reflect this, as a low frequency of teamwork was related to lower levels of collective efficacy than a moderate or high frequency of team experience.

A pattern was less easy to elucidate for the effect of the level of frequency of previous teamwork experience on errors, and most of the effects were negligible. High frequencies of past team experience were related to low Transfer errors, but the same was not true for Identify, Acknowledge, or Assess errors, in which high frequencies of past team experience resulted in the highest average errors. It is proposed that this lack of pattern is due to an interaction of the participant backgrounds and the SwS task. Most participants reported being involved in teams (N = 107, 96%), but individuals with military experience were specifically excluded from the study, meaning that these teams were not of the same type as the team required by the SwS task. When experience is not directly transferrable, it is less likely to help task and team performance (Orvis, Belanich, & Mullin, 2005). Therefore, it is likely that past team experience may affect communication, coordination, and task actions differently.

Familiarity with role and teammates

Lastly, familiarity with role was found to influence collective efficacy, such that collective efficacy tended to increase as participants gained experience in their role. The effect of the difference was small from Trial 2 to 3, which is not surprising, since prior evaluations of the learning effects showed that the largest difficulty drop off was from Trial 1 to Trial 2 (Ouverson, et al., 2018). It is worthwhile to note that Collective Efficacy was relatively high, even though performance was low, overall. This could indicate that a better scoring mechanism should be developed for the task, or it could be an example of the Dunning-Kruger effect, in which participants are underperforming but believe themselves to be performing fairly well (Dunning & David, 1999). Indeed, previous research aligned with this finding, with a reduction of this effect only for team-level feedback (Ostrander et al., 2019).

Familiarity with teammates was not found to have an effect on either individual or team performance; however, there was a moderately-sized trend in the Identify action for individuals who were totally familiar vs not familiar and a large-sized trend for those who were partially familiar vs not familiar. Participants who had partial familiarity with their teammates tended to have fewer Identify errors that those who were either fully or not at all familiar with their teammates prior to the team activity. Additionally, participants who had no familiarity tended to have fewer missed Transfers than those with partial familiarity. Neither of these trends was statistically significant, but the effect sizes for the identify trends were fairly strong, indicating that sample size may have played a role. More investigation is needed to understand why familiarity impacts these coordination actions and under what circumstances.

Limitations

Lastly, it is worth noting a few limitations of the study. As with any study, the ecological validity of the sample is called into question, and this study is no exception. On the one hand, the individuals being trained on the SwS task needed to be brought up to speed on military terminology and had no experience with the task or anything adjacent to it, which could make the learning experience more difficult, potentially beyond the scope of the tutor's abilities. In the present study, mitigation for this limitation was via extensive training prior to the four trials. On the other hand, the users are in a total naïve state, so if the tutor can train these participants how to complete the task successfully, users with prior experience should also be trained effectively by the system.

Relatedly, teamwork has been recognized as a core competency for college graduates (National Academies of Sciences, Engineering, and Medicine, 2017); therefore, college students can be expected to have been exposed to a great deal of teamwork experience. In addition to sheer exposure (nearly all students in the sample reported engaging in teamwork), affinity for teamwork among participants was high, as 88% of the sample reported that they enjoyed teamwork. Since recruitment materials specified that the study involved some degree of teamwork, this sample is expected to have included some degree of self-selection.

In order to reduce the volume of feedback that is given in the public feedback condition, Transfer actions were only given feedback in the private feedback condition. This was something that was deemed important when this legacy of ITTSs was being designed, starting with STT. This means that the Transfer action's response to feedback is automatically different in the public feedback condition, but that does not mean those results are not useful to further ITTS development.

Finally, there is little validation of the claims that Acknowledge actions are most similar to communication or that Transfer and Identify actions are coordination events. These ideas could be useful to future ITTSs, if true, and it would be useful to know how to categorize other user actions, just as others have categorized actions in face-to-face teams (Rosen et al., 2011; Salas et al., 2007). Future research could be directed to create a taxonomy of team skills relevant to ITTS-based training.

CHAPTER 7. GENERAL DISCUSSION & CONCLUSIONS

When considering all of the results of this study (Table 32), there are several strands to tie together. The following section offers a discussion of the conclusions for each of the components of the critical considerations framework used in this study: Cognition, Coordination, Communication, and Cooperation.

Table 32. A round-up of all hypotheses tested in this thesis and whether they were accepted or rejected, or if the result was partially supported.

	Alternative Hypothesis	Supported?
H1A:	For the participants who experience more than one role, Team Task Awareness will be higher.	Rejected
H1B:	For primary spotters who experience the sniper role, Sniper Goal Awareness will be higher.	*Partial Support*
H1C:	For the participants who experience more than one role, Shared Role Awareness will be higher.	Rejected
H2:	Participants who are familiar with at least one teammate will have fewer Acknowledgment errors.	*Partial Support*
H3:	Persons who work in teams more often score higher on Shared Role Awareness.	Rejected
H4:	In Trials 1 and 4, Acknowledge errors for participants will be higher than in Trials 2 and 3.	*Partial Support*
H5:	Public feedback will result in higher Shared Role Awareness than private feedback.	Rejected
H6:	Public feedback will result in lower Acknowledgment errors.	Rejected
H7A:	Teams receiving private feedback will perform better than teams receiving public feedback	*Partial Support*
H7B:	Private feedback will result in higher individual performance than Public feedback	*Partial Support*
H8:	Participants who use or find the feedback more helpful will have higher individual performance than those who do not use the feedback or do not find it as helpful	*Partial Support*
H9:	Persons who play a higher proportion co-op video games and play video games more often will have higher overall collective efficacy	Rejected
H10:	Participants with more frequent video game experience will have higher individual performance than those with less experience	Rejected
H11:	Participants with more frequent team experience will have a higher overall collective efficacy	*Partial Support*

Table 32. (continued)

Alternative Hypothesis		Supported?
H12:	Participants with more frequent team experience will have higher individual performance than those with less experience	*Partial Support*
H13:	Participants with experience in their role will have higher individual performance than participants with partial naïveté	*Partial Support*
H14A:	Participants who are more familiar with their teammates will have higher individual performance than those who are less familiar	*Partial Support*
H14B:	Teams with members who are fully familiar with one-another will perform better than teams with no and partial familiarity	Rejected
H15:	Collective efficacy will increase across Trials 1 through 4	**Full Support**

Cognition

Cognition was defined as each team member's understanding of the team experience. From the perspective of a participant in this experiment, this was the combination of an understanding of his or her primary task's responsibilities and goals, his or her teammate's responsibilities and goals, and the whole team coordination to fulfill those requirements. Specifically, the effects of feedback, experience, and familiarity on shared SA of the participants were explored within the journal manuscript, and these results add to this thesis in important ways.

Overall, shared SA was not shown to be very influenced by role familiarity or prior teamwork experience, which was an unexpected result when considering the literature. Previous work had shown that shared SA is improved by familiarity with required tasks and with team familiarity (MacMillan et al., 2002), but the contradiction revealed in this work may be more telling of the tutor than of the understanding of teamwork. However, teamwork experience, more broadly, had not been examined in relationship to shared SA development. In the case of broad experience with teamwork, while the skills in acquiring shared SA may be formed, there was no conclusive evidence that prior team experience impacts current development of shared SA (Sætrevik & Eid, 2014). On the other-hand, Shared SA, as it related to understanding the goals of the team, was shown to be marginally affected by role experience. Specifically, spotters who did not switch roles scored marginally lower on Sniper Goal Awareness.

Feedback privacy did not have a significant impact on shared SA. It was expected that since cross-training had been shown to be effective in fostering the development of shared SA (MacMillan et al., 2002; Volpe, Cannon-Bowers, Salas, & Spector, 2006), experience with feedback for other roles through publicly-presented feedback would also increase shared SA. However, no such relationship was shown in these data.

Coordination

Coordination was defined as the conversion of team-member resources into team-level outcomes. For participants in this experiment, coordination was realized through the handoffs of OPFOR from one spotter to the other as those OPFOR moved across zones and from the receiving spotter to the sniper after the OPFOR entered the new zone. In some sense, this refers to task performance, but these tasks were set up to require teamwork on the level of coordination in order for the tasks to be completed successfully (Gilbert et al., 2017).

This work revealed that coordination was impacted by feedback privacy, where private feedback was shown to positively impact coordination performance. Self-reported feedback use also had an effect on coordination, where coordination errors were higher for participants who reported not using the feedback compared to those who used the feedback selectively, either reporting ignoring it or noting it as distracting.

Additionally, information about role and teammate familiarity as well as experience with teams and their nuanced impact on performance was gleaned. Familiarity with teammates did not statistically significantly affect team performance. On coordination (individual performance), there was a moderately-sized effect for individuals who were totally familiar vs not familiar and a large-sized effect for those who were partially familiar vs not familiar. Neither of these effects were statistically significant, but the effect sizes for the identify trends were fairly strong, indicating that this lack of effect may be due to sample size.

Communication

Communication was defined as the information exchange among team members which guides teams to a common goal and a common understanding of that goal. For participants in this experiment, communication was realized both through verbal indications of the actions of targets within the game environment and through key-presses which alerted the tutor to player actions. Again, only the acknowledge key-press, which was deemed purely a communication task, was analyzed as communication. No verbal data was examined in this thesis.

The effects of feedback, experience, and familiarity on communication of the participants were explored within the journal manuscript, and these results add to this thesis in important ways. Communication was not affected by teammate familiarity. Previous research had examined quantity of communication for familiar and non-familiar teams and its effect on performance (Marlow et al., 2016), but not quality of communication in relationship to teammate familiarity. From the result revealed in this work, it is possible that while familiarity may make less communication necessary, it does not make the communication align with training expectations, despite the presence of feedback. It could also be that the communication measured within this study was extremely formulaic and structured, thereby overriding any impact of familiarity.

Additionally, feedback privacy did not have a statistically significant impact on communication. While public feedback was expected to lead to better communication performance (Alsharo et al., 2017; Cramton et al., 2007; Sætrevik & Eid, 2014; Windeler et al., 2015), private feedback was shown to result in better performance with a small effect size. This could be indicative of the effect of the tutor's communicative presence on human-human interactions, since the tutor had to vie for participant attention in order to give feedback. While steps were taken to limit the feedback volume differences between private and public feedback-receiving teams, this could be evidence that more should be done to mitigate this difference. However, this result could also signify that private feedback is better for performance even on more interpersonal skills, such as communication.

Communication was shown to be impacted by the trial from which the data was collected. Specifically, role naïveté put a strain on communication when that role was fresh, resulting in less accurate communication during those trials (in this case, Trial 1 and Trial 4).

Cooperation

Cooperation was defined as the beliefs of the members of the team which motivate teamwork behavior. For participants in this experiment, cooperation was measured via collective efficacy. There were impacts of task experience, frequency of previous team experience, and frequency of video-game specific team experience on collective efficacy.

Collective efficacy, or cooperation, was impacted by role experience. As participants moved through the experiment, they felt more confidence in their ability to perform as a team.

Additionally, cooperation beliefs were impacted by past team experience. For individuals who had a high-frequency history of teamwork (more than once per week), collective efficacy tended to be higher than for those with low past frequencies of experience. This trend was not statistically significant, and the effect sizes of the difference was small.

While not in alignment with the original hypothesis, cooperative video game experience had an interesting effect on collective efficacy. Only participants who played few video games overall showed differences in collective efficacy, depending on how many of those video games involved teamwork. A person who did not play video games very often could be expected to have less skill in the roles of that game, and as such, he or she reported lower levels of collective efficacy. During a cooperative video game, lower skill levels could amplify feelings of incompetence, as team members weigh-in on personal performance and team scores.

Future Work

Altogether this thesis offers insight into where future research is needed. ITTSs are arguably at the moment in which ITSs found themselves in twenty to thirty years ago. Few ITTSs have been developed and validated.

Future work should be focused on elaborating a number of the results of this thesis. First, the impact of an ITTS on the development of shared SA and shared mental models should be further explored. The scales used to evaluate this construct need to be re-evaluated and more systematically developed so as to properly assess the participants' understanding of the team's roles and the task. There is also room to tease apart the match between the participant's mental model and the mental model that is trained at the beginning of the experiment, either by coding the answers to the questions in a way that accounts for backing-up behaviors or by giving the quiz more frequently to observe the way the mental models shift over time.

Second, while evaluating the communication actions recorded by keypress as per request of the experimenters was a useful method for evaluating the effects observed by the tutor, this is not the only way to evaluate communication. The content of the spoken communications between teammates, the vocal tone, and the frequency of interruptions were not evaluated in this work. In the future, the recordings of participant verbal interactions should be parsed and evaluated to give a fuller picture. Another solution would be to incorporate sociometric badges that evaluate tone, frequency, and timing of utterances (Pentland, 2012).

Third, the nuanced effects of feedback privacy on individual performance, team performance, and communication highlight an important future step for ITTS research: truly adaptive tutoring. While the SwS tutor used evaluations of performance on predetermined behavioral markers to give feedback adapted to each individual's unique performance and only after a certain number of actions at a level outside of the user's current performance level, more sophisticated adaptation could be employed in the future. For example, the ITTS could monitor electrodermal activity (EDA) or electrical brain activity (via an EEG) to maintain a certain level of arousal or attention in its users, giving less feedback in times of particularly high-workload, or changing feedback content to foster attention and motivation, as in work done in human-agent teams by Szafir and Mutlu (2012).

Alternatively, feedback on specific types of actions could be adaptively triggered to account for user mastery or task-necessitation of that action. If the user has not missed a single Identify but he or she has recently begun to initiate more Identifies, this might be an indication of backing-up behavior rather than individual error. Further, if the tutor notices a certain action is less important to reaching team outcomes, the tutor could give less feedback on that behavior, focusing more attention on the important actions. Lastly, the tutor could adapt the frequency and choice of feedback based on behavioral changes noted over time by the teammates, taking a more complex performance history into account.

Feedback use should also be considered in future work. A combination of self-report measures or interviews and eye-tracking data would be the most informative for any future study hoping to understand the effect of feedback on performance. It is also advised that a control condition with no feedback be included. As mentioned, such a condition was not included here to reduce the number of manipulations and, therefore, the required size of the sample. However, this severely limits the conclusions that can be made about the performance of the tutor and its impacts on participant performance in the SwS task. In the present experiment, feedback use's tie to individual performance indicated the tutor had some effect, but without a control condition, the clarity of the relationship is muddled at best.

The work described here introduced a three-person iteration of a scalable team tutoring task and evaluated the effects of that tutor on teamwork components. This may serve as an exemplar for future work in developing real-time ITTS systems. Additionally, the results regarding the specific effects on different aspects of teamwork can be used help tailor the tutor to best meet the requirements set by the developing team.

More broadly, this research represents further exploration of team training, team tutoring, and socially capable tutors that seek to train team skills in addition to task skills. There is exciting progress being made toward the goal of virtual, software- or web-based team skill-training tutors, and future work should build on the basis of this and other work which has aimed to set the arrow of progress going in a productive direction.

REFERENCES

Alsharo, M., Gregg, D., & Ramirez, R. (2017). Virtual team effectiveness: The role of knowledge sharing and trust. *Information and Management*, *54*(4), 479–490. https://doi.org/10.1016/j.im.2016.10.005

Azevedo, R., Witherspoon, A., Chauncey, A., Burkett, C., & Fike, A. (2009). MetaTutor: A MetaCognitive Tool for Enhancing Self-Regulated Learning. *Cognitive and Metacognitive Educational Systems: Papers from the AAAI Fall Symposium*, 14–19. https://doi.org/10.1145/3012071.3012075

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147. Retrieved from https://pdfs.semanticscholar.org/8bee/c556fe7a650120544a99e9e063eb8fcd987b.pdf

Bandura, A. (1986). Self-Efficacy. In *Social foundations of thought and action: A social cognitive theory* (pp. 390–454). Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benefield, G. A., Shen, C., & Leavitt, A. (2016). Virtual Team Networks: How Group Social Capital Affects Team Success in a Massively Multiplayer Online Game. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 677–688. https://doi.org/10.1145/2818048.2819935

Billieux, J., Linden, M. Van Der, Achab, S., Khazaal, Y., Paraskevopoulos, L., Zullino, D., & Thorens, G. (2013). Why do you play World of Warcraft ? An in-depth exploration of self-reported motivations to play online and in-game behaviours in the virtual world of Azeroth. *Computers in Human Behavior*, *29*, 103–109. https://doi.org/10.1016/j.chb.2012.07.021

Bisbey, T. M., Reyes, D. L., Traylor, A. M., & Salas, E. (2019). Teams of psychologists helping teams: The evolution of the science of team training. *American Psychologist*, *74*(3), 278–289. https://doi.org/10.1037/amp0000419

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, *13*(6), 4–16. https://doi.org/10.3102/0013189X013006004

Bonner, D. (2017). *Sniper and Spotter Role Keyboard Use* [Digital Image].

Bonner, D., Gilbert, S., Dorneich, M., Burke, S., Walton, J., Ray, C., & Winer, E. (2014). Taxonomy of Teams, Team Tasks, and Tutors. *Proceedings of the 2nd Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium*, 1–9. Army Research Laboratory.

Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., … Holub, J. (2016). The Challenges of Building Intelligent Tutoring Systems for Teams. *Proceedings of the Human Factors & Ergonomics Society (HFES) 2016 Annual Meeting*, 1981–1985. https://doi.org/10.117/1541931213601451

Bonner, D., Gilbert, S., Winer, E., Dorneich, M., MacAllister, A., Kohl, A., … Sinatra, A. M. (2017). Military Team Training Utilizing GIFT. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 1–12. Orlando, FL.

Bradley, B. H., Klotz, A. C., Postlethwaite, B. E., & Brown, K. G. (2013). Ready to rumble: How team personality composition and task conflict interact to improve performance. *Journal of Applied Psychology*, *98*(2), 385–392. https://doi.org/10.1037/a0029845

Bradshaw, J. M., Feltovich, P. J., Johnson, M. J., Bunch, L., Breedy, M. R., Eskridge, T., … Uszok, A. (2008). Coordination in human-agent-robot teamwork. *2008 International Symposium on Collaborative Technologies and Systems, CTS'08*, *1*(850), 467–476. https://doi.org/10.1109/CTS.2008.4543966

Buche, C., Querrec, R., De Loor, P., & Chevaillier, P. (2004). Mascaret: Pedagogical multi-agents system for virtual environment for training. *Journal of Distance Education Technologies*, *2*(4), 41–61. Retrieved from https://hal.archives-ouvertes.fr/hal-00610857

Burke, C. S., Sottilare, R., Johnston, J., Sinatra, A., & Salas, E. (2017). *Towards a Scientifically Rooted Design Architecture of Team Process and Performance Modeling in Adaptive, Team-Based Intelligent Tutoring Systems*. Orlando, FL.

Cannon-Bowers, J., & Salas, E. (1998). Team Performance and Training in Complex Environments: Recent Findings from Applied Research. *Current Directions in Psychological Science*, *7*(3), 83–87. https://doi.org/10.1111/1467-8721.ep10773005

Carpenter, S., Fortune, J. L., Delugach, H. S., Etzkorn, L. H., Utley, D. R., Farrington, P. A., & Virani, S. (2008). Studying team shared mental models. *Proceedings of the 3rd International Conference on the Pragmatic Web: Innovating the Interactive Society*, 41–48. https://doi.org/10.1145/1479190.1479197

Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica*, *119*(2), 217–230. https://doi.org/10.1016/j.actpsy.2005.02.004

Chang, H. H., Hung, C., & Hsieh, H. (2014). Virtual teams: cultural adaptation, communication quality, and interpersonal trust. *Total Quality Management & Business Excellence*, *25*(11–12), 1318–1335. https://doi.org/10.1080/14783363.2012.704274

Charlier, S. D., Stewart, G. L., Greco, L. M., & Reeves, C. J. (2016). Emergent leadership in virtual teams: A multilevel investigation of individual communication and team dispersion antecedents. *Leadership Quarterly*, *27*(5), 745–764. https://doi.org/10.1016/j.leaqua.2016.05.002

Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012). Designing Persuasive Robots: How Robots Might Persuade People Using Vocal and Nonverbal Cues. *7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12)*, 293–300. https://doi.org/10.1145/2157689.2157798

Chou, H. W., Lin, Y. H., Chang, H. H., & Chuang, W. W. (2013). Transformational leadership and team performance: The mediating roles of cognitive trust and collective efficacy. *SAGE Open*, *3*(3), 1–10. https://doi.org/10.1177/2158244013497027

Cohen, J. (1988). The Effect Size Index: d. In *Statistical Power Analysis for the Behavioral Sciences* (Second). Retrieved from http://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal*, *19*(2), 237–248. https://doi.org/10.3102/00028312019002237

Colbry, Hurwitz, & Adair. (2014). Collaboration Theory. *Journal of Leadership Education*, *13*(4). https://doi.org/10.12806/V13/I4/C8

Corporation, V. (2019). Steam & Game Stats. Retrieved June 2, 2019, from https://store.steampowered.com/stats/

Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Craig, S. D., Hu, X., Graesser, A. C., Bargagliotti, A. E., Sterbinsky, A., Cheney, K. R., & Okwumabua, T. (2013). The impact of a technology-based mathematics after-school program using ALEKS on student's knowledge and behaviors. *Computers and Education*, *68*, 495–504. https://doi.org/10.1016/j.compedu.2013.06.010

Craighead, J. (2008). Distributed, game-based, intelligent tutoring systems - The next step in computer based training? *2008 International Symposium on Collaborative Technologies and Systems, CTS'08*, 247–256. https://doi.org/10.1109/CTS.2008.4543938

Cramton, C. D., Orvis, K. L., & Wilson, J. M. (2007). Situation invisibility and attribution in distributed collaborations. *Journal of Management*, *33*(4), 525–546. https://doi.org/10.1177/0149206307302549

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dabbish, L., Kraut, R., Patton, J., Heinz, H. J., & College, I. I. I. (2012). *Communication and Commitment in an Online Game Team*.

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A Multiple-Goal, Multilevel Model of Feedback Effects on the Regulation of

Individual and Team Performance. *Journal of Applied Psychology*, *89*(6), 1035–1056. https://doi.org/10.1037/0021-9010.89.6.1035

Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., & Hayes, C. C. (2012). Considering Etiquette in the Design of an Adaptive System. *Journal of Cognitive Engineering and Decision Making*, *6*(2), 243–265. https://doi.org/10.1177/1555343412441001

Driskell, J. E., Goodwin, G. F., Salas, E., & O'Shea, P. G. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics*, *10*(4), 249–271. https://doi.org/10.1037/1089-2699.10.4.249

Driskell, T., Salas, E., & Driskell, J. E. (2018). Teams in extreme environments: Alterations in team development and teamwork. *Human Resource Management Review*, *28*(4), 434–449. https://doi.org/10.1016/j.hrmr.2017.01.002

Dunning, J. K., & David. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Engineering, I. S. U. C. of. (2018). *The Surveillance Task with Sniper: A 3-Person ITTS Built with GIFT* [Online Video]. Retrieved from https://www.youtube.com/watch?v=IvGRcu7ialY&feature=youtu.be

Espevik, R., Johnsen, B. H., Eid, J., & Thayer, J. F. (2006). Shared mental models and operational effectiveness: Effects on performance and team processes in submarine attack teams. *Military Psychology*, *18*(SUPPL.). https://doi.org/10.1207/s15327876mp1803s_3

Falmagne, J.-C., Cosyn, E., Doignon, J.-P., & Thiéry, N. (2006). The Assessment of Knowledge, in Theory and in Practice. In *Formal Concept Analysis* (pp. 61–79). Berlin Heidelberg: Springer.

Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., … von Davier, A. (2017). Collaborative Problem Solving: Considerations for the National Assessment of Educational Progress. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*. https://doi.org/10.1145/2531602.2531690

Fowlkes, J., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *International Journal of Aviation Psychology*, *8*(3), 209–221. https://doi.org/10.1207/s15327108ijap0803_3

Freeman, G., & Wohn, D. Y. (2018). Understanding eSports Team Formation and Coordination. In *Computer Supported Cooperative Work: CSCW: An International Journal* (Vol. 27). https://doi.org/10.1007/s10606-017-9299-4

Freeman, J., Diedrich, F. J., Haimson, C., Diller, D. E., & Roberts, B. (2003). Behavioral Representations for Training Tactical Communication Skills. *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation*.

Freeman, J., & Zachary, W. (2018). Intelligent Tutoring for Team Training: Lessons Learned from US Military Research. In *Building Intelligent Tutoring Systems for Teams: What Matters* (Vol. 19, pp. 215–245). https://doi.org/10.1108/s1534-085620180000019013

Gallenkamp, J. V., Korsgaard, M. A., Welpe, I. M., Riedl, B. C., Picot, A., & Wigand, R. T. (2012). Conflict, value diversity, and performance in virtual teams. *18th Americas Conference on Information Systems 2012, AMCIS 2012*, *2*, 1137–1144. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84877911941&partnerID=tZOtx3y1

Geister, S., Konradt, U., & Hertel, G. (2006). Effects of Process Feedback on Motivation, Satisfaction, and Performance in Virtual Teams. *Small Group Research*, *37*(5), 459–489. https://doi.org/10.1177/1046496406292337

Gibbs, J. L., Sivunen, A., & Boyraz, M. (2017). Investigating the impacts of team type and design on virtual team processes. *Human Resource Management Review*, *27*(4), 590–603. https://doi.org/10.1016/j.hrmr.2016.12.006

Gilbert, S. B., Sinatra, A. M., Macallister, A., Kohl, A., Winer, E., Slavina, A., … Ouverson, K. M. (2016). Analyzing Team Training Data : Aspirations for a GIFT Data Analytics Engine. *Industrial and Manufacturing Systems Engineering Conference Proceedings and Posters*. Retrieved from http://lib.dr.iastate.edu/imse_conf/99

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., … Winer, E. (2017). Creating a Team Tutor Using GIFT. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-017-0151-2

Gilbert, S., Winer, E., Holub, J., Richardson, T., Dorneich, M., & Hoffman, M. (2015). Characteristics of a Multi-User Tutoring Architecture. *Proceedings of the 3rd Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)*. Orlando, FL: Army Research Laboratory.

Gorman, J. C., Cooke, N. J., Winner, J. L., Duran, J. L., Pedersen, H. K., & Taylor, A. R. (2007). Knowledge Training versus Process Training: The Effects of Training Protocol on Team Coordination and Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 51, pp. 382–386. https://doi.org/10.1177/154193120705100449

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Graesser, A. C., Hu, X., & Sottilare, R. (2017). Intelligent Tutoring Systems. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of* *the Learning Sciences* (pp. 246–255). New York: Routledge.

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent Tutoring Systems with Conversational Dialogue. *AI Magazine*, *22*(4), 39–52. https://doi.org/http://dx.doi.org/10.1609/aimag.v22i4.1591

Graesser, A., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*(4), 612–618. https://doi.org/10.1109/TE.2005.856149

Haines, R. (2014). Group development in virtual teams: An experimental reexamination. *Computers in Human Behavior*, *39*, 213–222. https://doi.org/10.1016/j.chb.2014.07.019

He, F., Paul, S., & Dennis, A. R. (2018). Group Atmosphere, Shared Understanding, and Team Conflict in Short Duration Virtual Teams. *Hawaii International Conference on System Sciences*, *51*, 362–370.

Hirsch, P. L., & Mckenna, A. F. (2008). Using Reflection to Promote Teamwork Understanding in Engineering Design Education. *International Journal of Engineering Education*, *24*(2), 377–385.

Hoch, J. E., & Dulebohn, J. H. (2017). Team personality composition, emergent leadership and shared leadership in virtual teams: A theoretical framework. *Human Resource Management Review*, *27*(4), 678–693. https://doi.org/10.1016/j.hrmr.2016.12.012

Hoch, J. E., & Kozlowski, S. W. J. (2014). Leading virtual teams: Hierarchical leadership, structural supports, and shared team leadership. *Journal of Applied Psychology*, *99*(3). https://doi.org/10.1037/a0030264

Katz-Navon, T. Y., & Erez, M. (2005). When collective- And self-efficacy affect team performance the role of task interdependence. *Small Group Research*, *36*(4), 437–465. https://doi.org/10.1177/1046496405275233

Koedinger, K. R., Brunskill, E., Baker, R. S. J. D., Mclaughlin, E. A., & Stamper, J. (2013). New potentials for data-driven inteligent tutoring system development and optimization. *AI Magazine*, *34*(3), 27–41. https://doi.org/http://dx.doi.org/10.1609/aimag.v34i3.2484

Koedinger, K. R., Matsuda, N., Maclellan, C. J., & McLaughlin, E. A. (2015). Methods for evaluating simulated learners: Examples from SimStudent. *CEUR Workshop Proceedings*, *1432*, 45–54.

Kou, Y., & Gui, X. (2014). Playing with strangers: understanding temporary teams in league of legends. *CHI PLAY ’14: Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play*, 161–169. https://doi.org/10.1145/2658537.2658538

Kruijff, G. J., Janíček, M., & Zender, H. (2012). Situated communication for joint activity in human-robot teams. *IEEE Intelligent Systems*, *27*(2), 27–35.

https://doi.org/10.1109/MIS.2012.8

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. *Intelligent Tutoring Systems*, *6094 LNCS*(PART 1), 156–164. https://doi.org/10.1007/978-3-642-13388-6_20

Lane, H. C., Core, M. G., Gomboc, D., Karnavat, A., & Rosenberg, M. (2007). Intelligent Tutoring for Interpersonal and Intercultural Skills Report Documentation Page. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 1–11.

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Retrieved from https://cran.r-project.org/package=emmeans

Locke, E. A., & Latham, G. P. (2015). Breaking the rules: A historical overview of goal-setting theory. *Advances in Motivation Science*, *2*(1973), 99–126. https://doi.org/http://dx.doi.org/10.1016/bs.adms.2015.05.001

Lohani, M., Stokes, C., Mccoy, M., Bailey, C. A., & Rivers, S. E. (2016). Social interaction moderates human-robot trust-reliance relationship and improves stress coping. *ACM/IEEE International Conference on Human-Robot Interaction*, *2016–April*, 471–472. https://doi.org/10.1109/HRI.2016.7451811

MacMillan, J., Paley, M. J., Levchuk, Y. N., Entin, E. E., Serfaty, D., & Freeman, J. T. (2002). Designing the best team for the task: Optimal organizational structures for military missions. In M. McNeese, E. Salas, & M. Endsley (Eds.), *New trends in cooperative activities: System dynamics in complex settings* (pp. 284–299). San Diego, CA: Human Factors and Ergonomics Society Press.

Marlow, S. L., Lacerenza, C. N., & Salas, E. (2016). Communication in virtual teams: a conceptual framework and research agenda. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2016.12.005

Mason, W., & Clauset, A. (2013). Friends FTW! friendship and competition in halo: reach. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*, 375–386. https://doi.org/10.1145/2441776.2441820

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In *In R. A. Guzzo & E. Salas (Eds.), Team effectiveness and decision making in organizations. San Francisco: Jossey-Bass* (pp. 9–45).

Mumm, J., & Mutlu, B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior*, *27*(5), 1643–1650. https://doi.org/10.1016/j.chb.2011.02.002

Nardi, B. (2005). Beyond bandwidth: Dimensions of connection in interpersonal communication. *Computer Supported Cooperative Work: CSCW: An International Journal*, *14*(2), 91–130. https://doi.org/10.1007/s10606-004-8127-9

National Academies of Sciences Engineering and Medicine, E. (2017). Intra- and Interpersonal Competencies as College Outcomes. In J. Herman & M. Hilton (Eds.), *Supporting Students' Success: The Role of Assessment of Intrapersonal and Interpersonal Competencies* (pp. 153–180). https://doi.org/https://doi.org/10.17226/24697

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427–469. https://doi.org/10.1007/s40593-014-0029-5

Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., & Cassell, J. (2012). "Oh dear stacy!": social interaction, elaboration, and learning with teachable agents. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*, 39–48. Retrieved from http://dl.acm.org/citation.cfm?id=2207676.2207684

Oren, M. A., & Gilbert, S. B. (2010). Interfaces for Communication Intervention: Utilizing social theory to support interdisciplinary design communication. *The Design Communication Workshop at Design, Creativity, and Cognition (DCC)*. Stuttgart.

Oren, M. A., & Gilbert, S. B. (2011). Framework for measuring social affinity for CSCW software. *CHI EA '11: CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 1387–1392. https://doi.org/10.1145/1979742.1979779

Orvis, K. L., Ruark, G., Engel, K. L., & Ratwani, K. L. (2010). Training Leaders on Emotion Management Skills : Challenges in Designing a Blended Learning Program. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, 2284–2288. https://doi.org/10.1518/10711810X12829370266329

Orvis, K., Orvis, K. L., Belanich, J., & Mullin, L. N. (2005). The Influence of Trainee Gaming Experience and Computer Self-Efficacy on Learner Outcomes of Videogame-Based Learning Environments. In *Army Technical Report* (Vol. 1164). Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a437016.pdf

Ostrander, A. (2019). *Inferring Team Metrics from Incomplete Information*. Iowa State University.

Ostrander, A., Bonner, D., Walton, J., Slavina, A., Ouverson, K. M., Kohl, A., … Winer, E. (2019). Evaluation of an Intelligent Team Tutoring System for a Collaborative Two-Person Problem: The Surveillance Team Tutor. *Computers in Human Behavior*.

Ouverson, K., Ostrander, A., MacAllister, A., Kohl, A., Walton, J., Gilbert, S., … Sinatra, A. M. (2018). Team Training for Enemy Identification Using an Intelligent Tutoring System. *Proceedings of Interservice/Industry Training, Simulation, and Education*

Conference (I/ITSEC), (18291), 1–14.

Ouverson, K., Pena, M. I., Walton, J., Gilbert, S., & Dorneich, M. (2018). Applying Human-Agent Team Concepts to the Design of Intelligent Team Tutoring Systems. *Intelligent Tutoring Systems*, 457–459. Retrieved from http://link.springer.com/10.1007/978-3-319-91464-0

Paul, R., Drake, J. R., & Liang, H. (2016). Global Virtual Team Performance: The Effect of Coordination Effectiveness, Trust, and Team Cohesion. *IEEE Transactions on Professional Communication*, *59*(3). https://doi.org/10.1109/TPC.2016.2583319

Paul, S., & Ray, S. (2014). Group atmosphere, shared understanding, and perceived conflict in virtual teams: Findings from an experiment. *20th Americas Conference on Information Systems, AMCIS 2014*, (2001), 1–10. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84905980723&partnerID=tZOtx3y1

Peñarroja, V., Orengo, V., Zornoza, A., Sánchez, J., & Ripoll, P. (2015). How team feedback and team trust influence information processing and learning in virtual teams: A moderated mediation model. *Computers in Human Behavior, 48*, 9–16. https://doi.org/10.1016/j.chb.2015.01.034

Pentland, A. S. (2012, April). The new Science of Building Great Teams: Analytics for Success. *Harvard Business Review*, (April), 1–11. https://doi.org/citeulike-article-id:10606943

Person, N., & Graesser, A. C. (2002). Human or Computer? AutoTutor in a Bystander Turing Test. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Proceedings of the 6th International Conference on Intelligent Tutoring Systems (ITS '02)* (pp. 821–830). https://doi.org/10.1007/3-540-47987-2_82

Peters, L. M., & Manz, C. C. (2007). Identifying antecedents of virtual team collaboration. *Team Performance Management: An International Journal*, *13*(3/4), 117–129. https://doi.org/10.1108/13527590710759865

Pinjani, P., & Palvia, P. (2013). Trust and knowledge sharing in diverse global virtual teams. *Information & Management*, *50*(4), 144–153. https://doi.org/10.1016/j.im.2012.10.002

Porter, C. O. L. H., Gogus, C. I., & Yu, R. C. F. (2011). Does backing up behavior explain the efficacy-performance relationship in teams? *Small Group Research*, *42*(4), 458–474. https://doi.org/10.1177/1046496410390964

Price, M., Mudrick, N., Taub, M., & Azevedo, R. (2018). The Role of Negative Emotion Regulation on Self-Regulated Learning with MetaTutor. *14th International Conference on Intelligent Tutoring Systems (ITS 2018)*.

Reagans, R., Argote, L., & Brooks, D. (2005). Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and

Knowing How to Work Together. *Management Science*, *51*(6), 869–881. https://doi.org/10.1287/mnsc.1050.0366

Rentsch, J. R., Heffner, T. S., & Duffy, L. T. (1994). What you know is what you get from experience : Team experience related to teamwork schemas. *Group & Organization Management*, *19*(4), 450–474.

Rentsch, J. R., & Klimoski, R. J. (2001). Why do “great minds” think alike? *Journal of Organizational Behavior*, (2), 107–120.

Rickel, J., & Johnson, W. L. (1999). Virtual humans for team training in virtual reality. *Proceedings of the 9th International Conference on Computers in Education (ICCE)*, (April 1999), 3–11.

Riggio, R. E., & Lee, J. (2007). Emotional and interpersonal competencies and leader development. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2007.08.008

Rosen, M. A., Bedwell, W. L., Wildman, J. L., Fritzsche, B. A., Salas, E., & Burke, C. S. (2011). Managing adaptive performance in teams: Guiding principles and behavioral markers for measurement. *Human Resource Management Review*, *21*(2), 107–122. https://doi.org/10.1016/j.hrmr.2010.09.003

Rudawska, A. (2017). Students’ Team Project Experiences and Their Attitudes Towards Teamwork. *Journal of Management and Business Administration. Central Europe*, *25*(1), 78–97. https://doi.org/10.7206/jmba.ce.2450-7814.190

Sabourin, J., Mott, B., & Lester, J. C. (2011). Modeling learner affect with theoretically grounded dynamic Bayesian networks. *International Conference on Affective Computing and Intelligent Interaction*, 286–295. Berlin Heidelberg: Springer.

Sætrevik, B., & Eid, J. (2014). The “similarity Index” as an indicator of shared mental models and situation awareness in field studies. *Journal of Cognitive Engineering and Decision Making*, *8*(2), 119–136. https://doi.org/10.1177/1555343413514585

Salas, E., Cooke, N., & Rosen, M. (2008). On teams, teamwork, and team performance: discoveries and developments. *Human Factors*, *50*(3), 540–547. https://doi.org/10.1518/001872008X288457.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). https://doi.org/10.1007/s13398-014-0173-7.2

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation Space and Environmental Medicine*, *78*(5), B77-85.

Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2009). Performance Measurement in Simulation-Based Training. *Simulation & Gaming*, *40*(3), 328–376. https://doi.org/10.1177/1046878108326734

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and Improving Teamwork in Organizations: A Scientifically Based Practical Guide. *Human Resource Management*, *54*(4), 599–622. https://doi.org/10.1002/hrm.21628

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, *13*(1), 89–99. https://doi.org/10.1016/0883-0355(89)90018-9

Sinatra, A. M., Gilbert, S. B., Dorneich, M. C., Winer, E., Ostrander, A., & Ouverson, K. M. (2018). Considerations for Dealing with Real-Time Communications in an Intelligent Team Tutoring System Experiment. *Proceedings of Workshop on Assessment and Intervention during Team Tutoring at AIED2018*, 28–35. CEUR-WS.org.

Smith-Jentsch, K. A., Sierra, M. J., Weaver, S. J., Bedwell, W. L., Dietz, A. S., Carter-Berenson, D., … Salas, E. (2015). *Training "The Right Stuff": An Assessment of Team Training Needs For Long-Duration Spaceflight Crews*. Retrieved from https://ston.jsc.nasa.gov/collections/trs/_techrep/TM-2015-218589.pdf

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). *The Generalized Intelligent Framework for Tutoring (GIFT)*. https://doi.org/10.13140/2.1.1629.6003

Sottilare, R. A., Holden, H. K., Brawner, K. W., & Goldberg, B. S. (2011). Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. https://doi.org/10.1128/AEM.71.11.7253

Sottilare, R. A., Shawn Burke, C., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing Adaptive Instruction for Teams: a Meta-Analysis. *International Journal of Artificial Intelligence in Education*, *28*(2), 225–264. https://doi.org/10.1007/s40593-017-0146-z

Stajkovic, A. D., Lee, D., & Nyberg, A. J. (2009). Collective Efficacy, Group Potency, and Group Performance: Meta-Analyses of Their Relationships, and Test of a Mediation Model. *Journal of Applied Psychology*, *94*(3), 814–828. https://doi.org/10.1037/a0015659

Stillson, H., & Alsup, J. (2003). Smart ALEKS ... or Not? Teaching Basic Algebra using an Online Interactive Learning System. *Mathematics & Computer Education*, *37*(3), p329-340.

Stillson, H., & Nag, P. (2009). Aleks and Mathxl: Using Online Interactive Systems To Enhance a Remedial Algebra Course. *Mathematics and Computer Education*, *43*(3), 239–247. Retrieved from http://ezproxy.usherbrooke.ca/login?url=http://search.proquest.com/docview/235944433?accountid=13835

Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (2011). The Role of Shared Mental Models in Developing Team Situational Awareness: Implications for Training. In E. Salas & A. S. Dietz (Eds.), *Situational Awareness* (1st ed., pp. 287–318). https://doi.org/10.4324/9781315087924-18

Strobach, T., & Schubert, T. (2013). Positive consequences of action-video game experience on human cognition: Potential benefits on a societal level. In K. K. Mak (Ed.), *Epidemiology of Online Game Addiction* (Vol. 49, pp. 1–6). OMICS Group.

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size-or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Szafir, D., & Mutlu, B. (2012). Pay attention!: designing adaptive agents that monitor and improve user engagement. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI ’12*, 11–20. https://doi.org/10.1145/2207676.2207679

Tasa, K., Taggar, S., & Seijts, G. H. (2007). The development of collective efficacy in teams: A multilevel and longitudinal perspective. *Journal of Applied Psychology*, *92*(1), 17–27. https://doi.org/10.1037/0021-9010.92.1.17

Tchogovadze, G. G. (1986). Some steps towards intelligent computer tutoring systems. *Education and Computing*, *2*(4), 273–277. https://doi.org/10.1016/S0167-9287(86)91437-8

Thomas, D. M., & Bostrom, R. P. (2007). The Role of a Shared Mental Model of Collaboration Technology in Facilitating Knowledge Work in Virtual Teams. *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*, 37. https://doi.org/10.1109/HICSS.2007.552

Thompson, N. A. (2010). *Encyclopedia of Research Design* (N. Salkind, Ed.). https://doi.org/10.4135/9781412961288.n205

Timperley, H., & Hattie, J. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Tong, Y., Yang, X., & Teo, H. H. (2013). Spontaneous virtual teams: Improving organizational performance through information and communication technology. *Business Horizons*, *56*(3), 361–375. https://doi.org/10.1016/j.bushor.2013.01.003

Traum, D., Rickel, J., Gratch, J., & Marsella, S. (2003). Negotiation over tasks in hybrid human-agent teams for simulation-based training. *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems - AAMAS ’03*, (January), 441. https://doi.org/10.1145/860575.860646

Turel, O., & Zhang, Y. (Jenny). (2010). Does virtual team composition matter? Trait and problem-solving configuration effects on team performance. *Behaviour & Information Technology*, *29*(4), 363–375. https://doi.org/10.1080/01449291003752922

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Volpe, C. E., Cannon-Bowers, J. A., Salas, E., & Spector, P. E. (2006). The Impact of Cross-Training on Team Functioning: An Empirical Investigation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(1), 87–100. https://doi.org/10.1518/001872096778940741

Walton, J., Dorneich, M. C., Gilbert, S., Bonner, D., Winer, E., & Ray, C. (2014). Modality and Timing of Team Feedback: Implications for GIFT. *Proceedings of the 2nd Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium*, 1–8. Army Research Laboratory.

Walton, J., Gilbert, S., Winer, E., Dorneich, M., & Bonner, D. (2015). Evaluating Distributed Teams with the Team Multiple Errands Test. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 1–12. https://doi.org/10.1145/2685553.2699018

Walton, J., Ostrander, A., Ouverson, K. M., Gilbert, S. B., Dorneich, M. C., Winer, E., & Sinatra, A. M. (2018). Feedback Design Considerations for Intelligent Team Tutoring Systems. *Human Factors and Ergonomics Society's 2018 International Annual Meeting*. Philadelphia, PA.

Wiese, C. W., Shuffler, M. L., & Salas, E. (2015). Teamwork and Team Performance Measurement. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (Second Edi, Vol. 24). https://doi.org/10.1016/B978-0-08-097086-8.22017-5

Windeler, J. B., Robert, L. P., & Riemenschneider, C. K. (2015). E-profiles, Conflict, and Shared Understanding in Distributed Teams. *Journal of the Association for Information*, *16*(7), 608–645.

Yang, E., & Dorneich, M. (2018). Evaluating Human–Automation Etiquette Strategies to Mitigate User Frustration and Improve Learning in Affect-Aware Tutoring. *Applied Sciences*, *8*(6), 895. https://doi.org/10.3390/app8060895

Yang, E., & Dorneich, M. C. (2016). Evaluation of Etiquette Strategies to Adapt Feedback In Affect-Aware Tutoring. *Human Factors and Ergonomics Society's International Annual Meeting*, 393–397. https://doi.org/10.1177/1541931213601089

Yin, J., Miller, M. S., Ioerger, T. R., Yen, J., & Volz, R. a. (2000). A knowledge-based approach for designing intelligent team training systems. *Proceedings of the Fourth International Conference on Autonomous Agents - AGENTS '00*, 427–434. https://doi.org/10.1145/336595.337560

Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. (1999). The Advanced Embedded Training System (AETS): An Intelligent Embedded Tutoring System for Tactical Team Training. *International Journal of Artificial Intelligence in Education (IJAIED)*, *10*, 257–277.

APPENDIX A. ADDITIONAL TABLES

Table 33. Count and percentage of the number of participants who noticed the feedback in each trial, and to what degree they found it helpful.

Response	**Trial** T1	T2	T3	T4	Total
I ignored the feedback	*12* *10.8%*	*2* *1.8%*	*3* *2.7%*	*7* *6.6%*	*24* *5.5%*
No, it was actually distracting	*1* *0.9%*	*2* *1.8%*	*9* *8.1%*	*2* *1.9%*	*14* *3.2%*
No, it was not very helpful	*20* *18%*	*12* *10.8%*	*19* *17.1%*	*20* *18.9%*	*71* *16.2%*
Yes, it was somewhat helpful	*38* *34.2%*	*57* *51.4%*	*36* *32.4%*	*31* *29.2%*	*162* *36.9%*
Yes, it was very helpful	*7* *6.3%*	*17* *15.3%*	*30* *27%*	*28* *26.4%*	*82* *18.7%*
Did not notice feedback	*32* *28.8%*	*20* *18%*	*13* *11.7%*	*17* *15.3%*	*82* *18.7%*
Total	*111*	*111*	*111*	*106*	*439*

Table 34. Correlation Matrix of all variables measured at an individual level

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1 Familiarity Score	1.00							
2 Team Task Awareness	-0.19	1.00						
3 Sniper Goal Awareness	-0.04	0.21	1.00					
4 Shared Role Awareness	-0.12	0.23	0.43	1.00				
5 Video Game Experience	-0.19	0.17	0.12	0.14	1.00			
6 Co-op Video Game Frequency	-0.02	0.02	-0.01	-0.04	0.31	1.00		
7 Feedback Use	0.02	-0.00	0.16	0.22	-0.00	-0.12	1.00	
8 Collective Efficacy	0.04	0.08	-0.07	-0.02	0.12	0.03	-0.13	1.00
9 Team Experience	-0.17	0.05	-0.04	0.03	-0.07	0.04	0.10	0.09
10 Communication performance	-0.04	0.01	0.02	0.09	0.03	0.06	-0.11	-0.01
11 Communication Errors	0.12	-0.08	-0.09	-0.03	-0.18	0.15	-0.08	-0.26
12 Assess Errors	-0.04	0.03	-0.20	-0.04	-0.09	-0.28	-0.14	0.13
13 Identify Errors	0.12	-0.19	-0.04	-0.23	0.08	0.14	-0.13	0.00
14 Transfer Errors	0.00	-0.16	0.10	0.04	-0.13	0.00	0.09	0.13
15 Feedback Type	-0.08	0.14	0.11	0.14	-0.04	-0.09	-0.11	-0.01

Table 34. (continued)

	9	**10**	**11**	**12**	**13**	**14**	**15**
9 Team Experience	1.00						
10 Communication performance	-0.08	1.00					
11 Communication Errors	-0.04	0.41	1.00				
12 Assess Errors	0.07	-0.29	-0.21	1.00			
13 Identify Errors	-0.07	-0.27	-0.16	--	1.00		
14 Transfer Errors	0.01	-0.13	-0.10	--	0.09	1.00	
15 Feedback Type	-0.11	0.14	0.08	0.14	-0.07	-0.16	1.00

APPENDIX B. ADDITIONAL FIGURES

The normality of each model was evaluated by visual inspection of the residual plots and the quantile-quantile plots for the error term. These plots are included here. While the assumption of normality was not so clearly met for all models, they were considered "normal enough" and results were interpreted with that understanding.

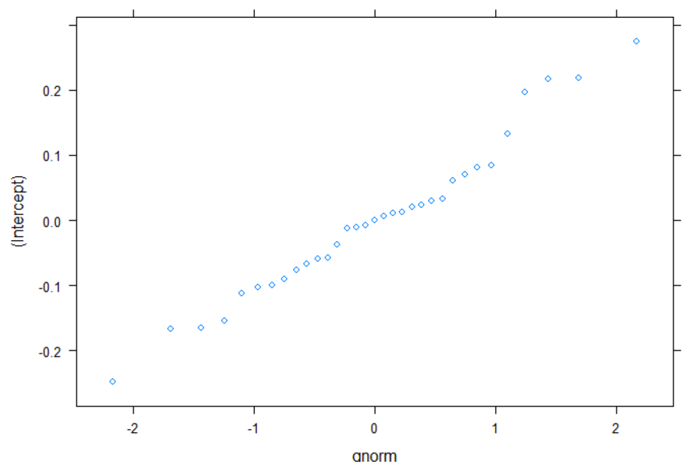

Figure 25. Quantile plot of the random effect of team for the Communication model.

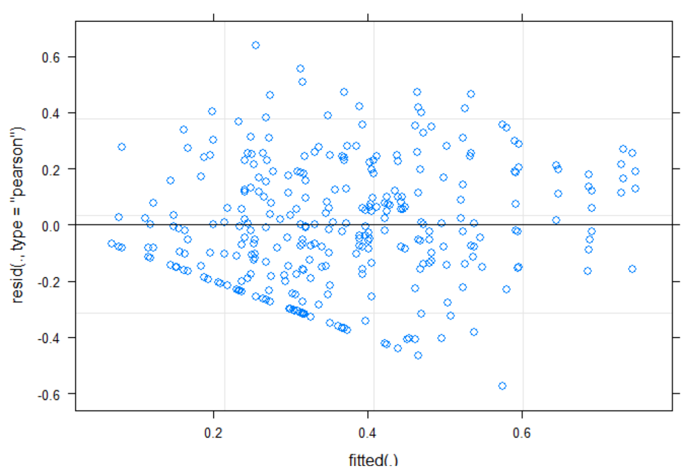

Figure 26. Scatterplot of the fitted Communication model against its residuals.

The model of the effect of teammate familiarity, feedback condition, and trial on Communication was evaluated for normality of errors (Figure 25) and the normality of the residuals (Figure 26). The model of the effect of the role switch on Sniper Goal Awareness was evaluated for normality of errors (Figure 27) and the normality of the residuals (Figure 28).

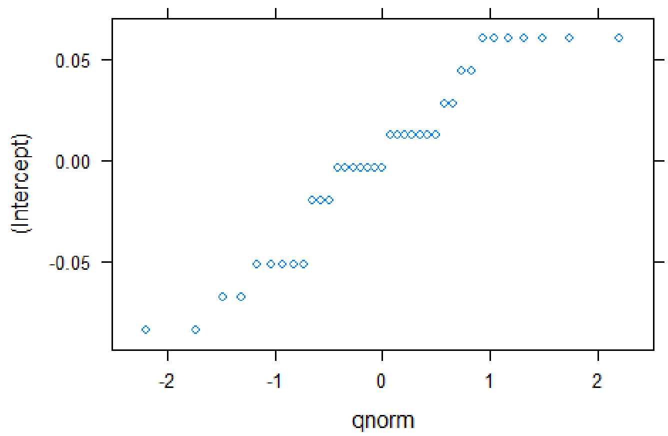

Figure 27. Quantile plot of the random effect of team for the Sniper Goal Awareness model.

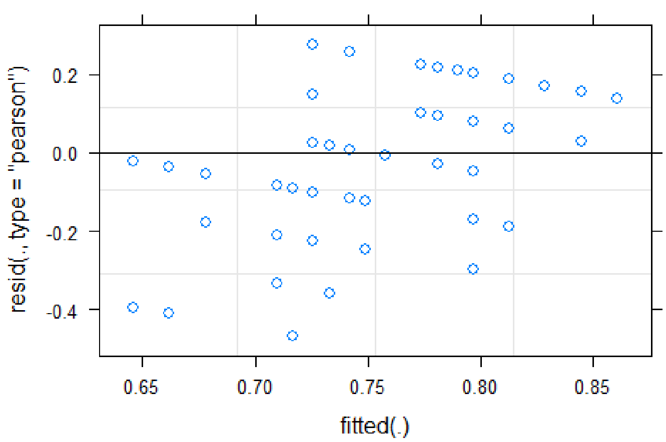

Figure 28. Scatterplot of the fitted Sniper Goal Awareness model against its residuals.

The model of the effect of the role switch, the feedback condition, and the frequency of prior team experience on Shared Role Awareness was evaluated for normality of errors (Figure 29) and the normality of the residuals (Figure 30).

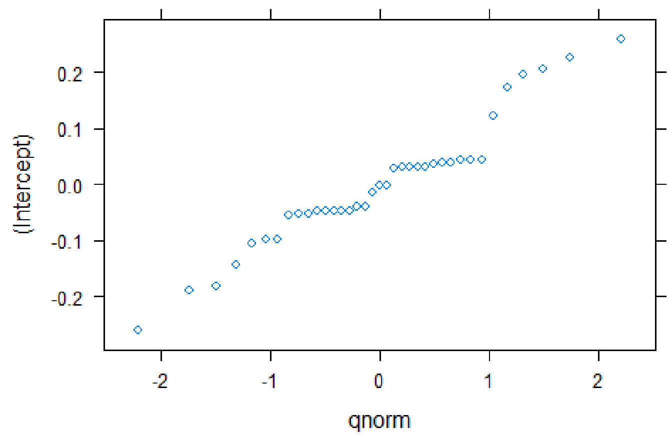

Figure 29. Quantile plot of the random effect of team for the Shared Role Awareness model.

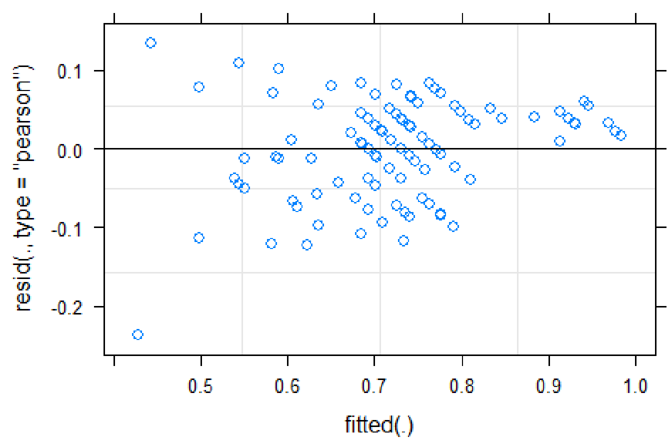

Figure 30. Scatterplot of the fitted Shared Role Awareness model against its residuals.

The model of the effect of the role switch on Team Task Awareness was evaluated for normality of errors (Figure 31) and the normality of the residuals (Figure 32).

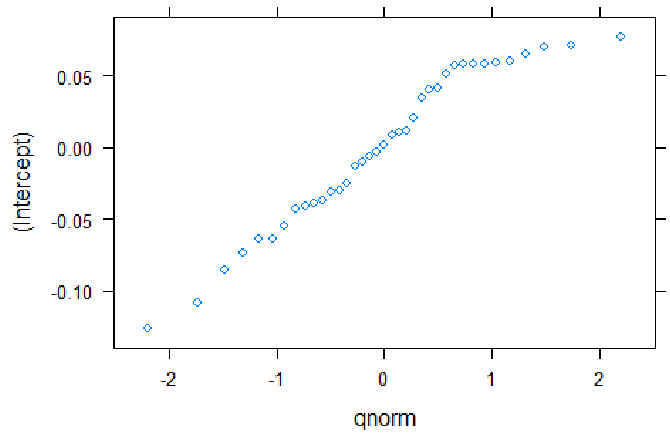

Figure 31. *Quantile plot of the random effect of team for the Team Task Awareness model.*

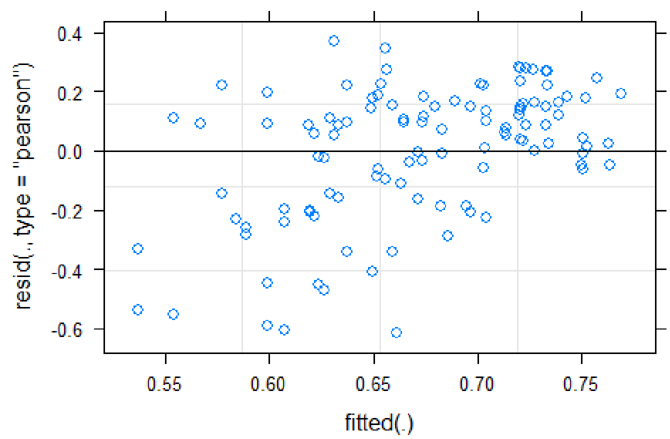

Figure 32. *Scatterplot of the fitted Team Task Awareness model against its residuals.*

The model of the effect of trial, the frequency of prior team experience, and the interaction of the amount of prior cooperative video game experience and the frequency of total prior video game experience on Collective Efficacy was evaluated for normality of errors (Figure 33) and the normality of the residuals (Figure 34).

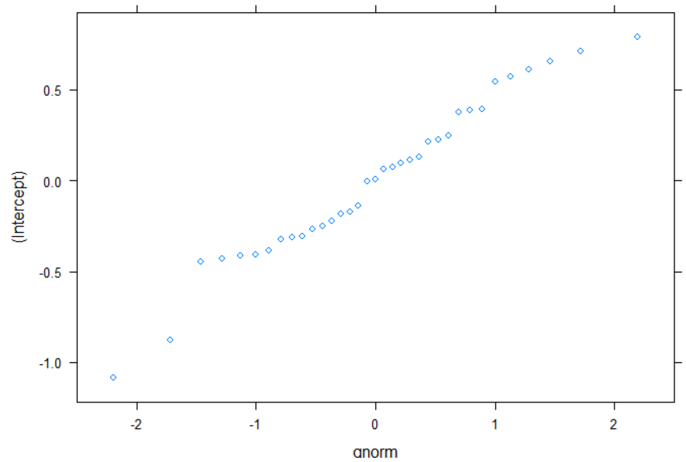

Figure 33. Quantile plot of the random effect of team for the Collective Efficacy model.

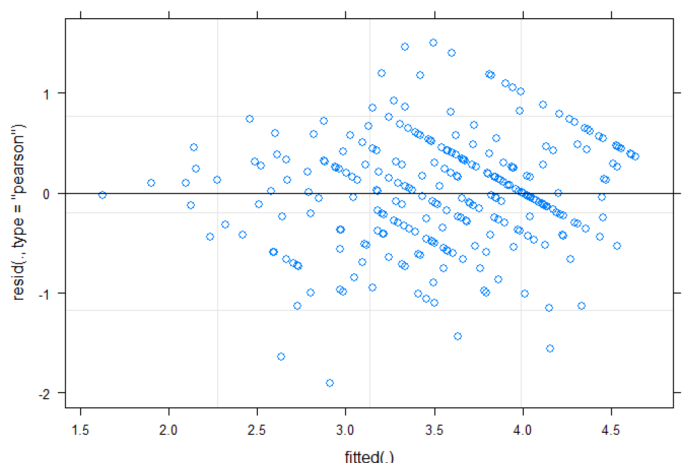

Figure 34. Scatterplot of the fitted Collective Efficacy model against its residuals.

The model of the effect of the feedback condition and the level of teammate familiarity on Team Performance was evaluated for normality of errors (Figure 35) and the normality of the residuals (Figure 36).

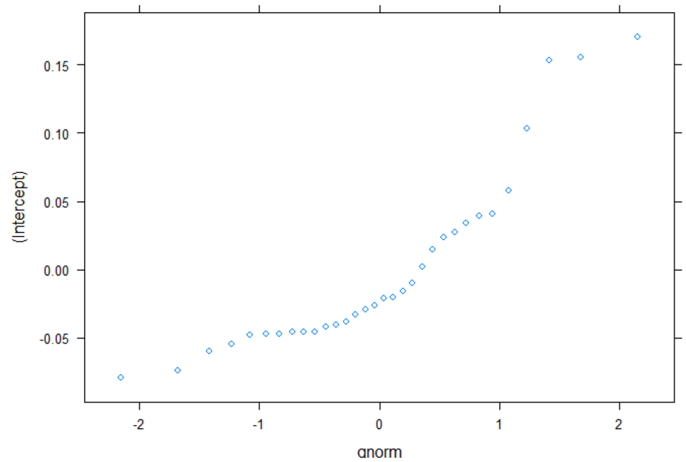

Figure 35. *Quantile plot of the random effect of team for the Team Performance model.*

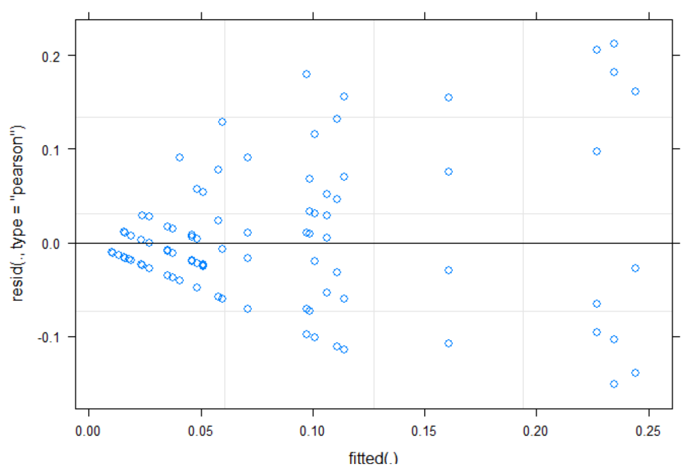

Figure 36. *Scatterplot of the fitted Team Performance model against its residuals.*

The model of the effect of the feedback condition, primary role, the frequency of prior video game experience, the frequency of prior video game experience, the level of self-reported feedback use, and the level of teammate familiarity on Transfer Errors was evaluated for normality of errors (Figure 37) and the normality of the residuals (Figure 38).

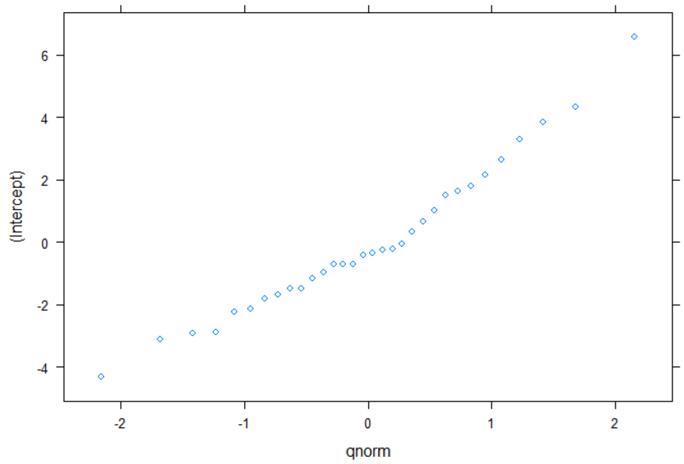

Figure 37. Quantile plot of the random effect of team for the Transfer Errors model.

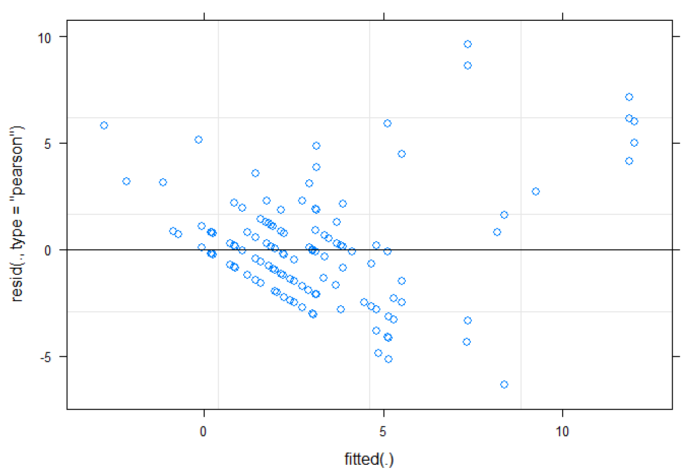

Figure 38. Scatterplot of the fitted Transfer Errors model against its residuals.

The model of the effect of the feedback condition, primary role, the frequency of prior video game experience, the frequency of prior video game experience, the level of self-reported feedback use, and the level of teammate familiarity on Acknowledge Errors was evaluated for normality of errors (Figure 39) and the normality of the residuals (Figure 40).

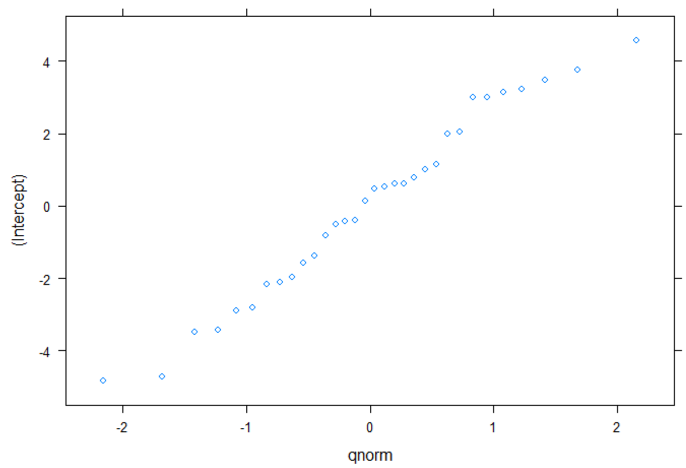

Figure 39. Quantile plot of the random effect of team for the Acknowledge Errors model.

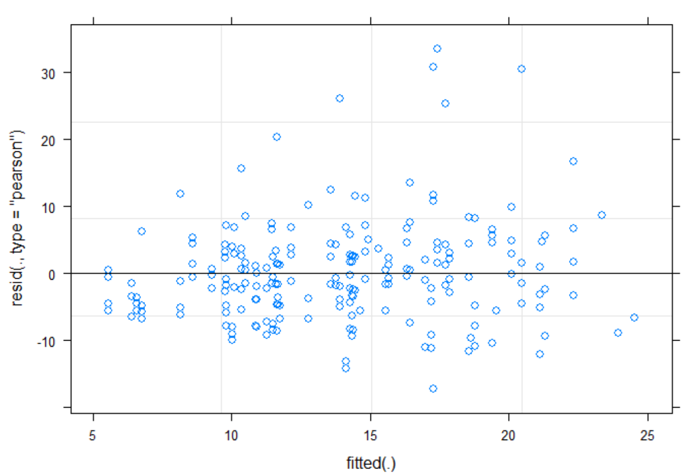

Figure 40. Scatterplot of the fitted Acknowledge Errors model against its residuals.

The model of the effect of the feedback condition, primary role, the frequency of prior video game experience, the frequency of prior video game experience, the level of self-reported feedback use, and the level of teammate familiarity on Identify Errors was evaluated for normality of errors (Figure 41) and the normality of the residuals (Figure 42).

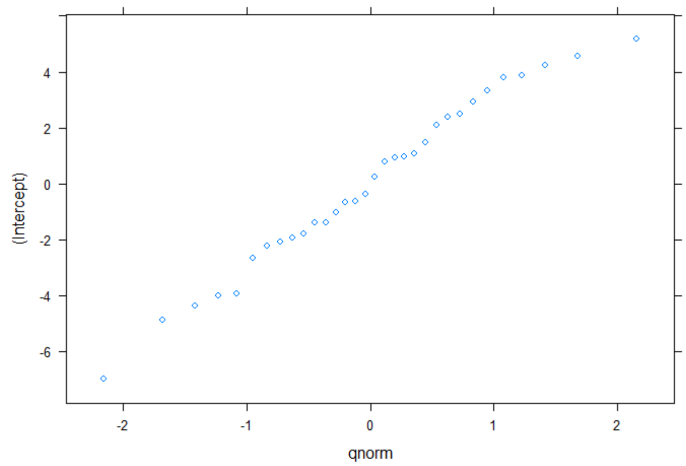

Figure 41. Quantile plot of the random effect of team for the Identify Errors model.

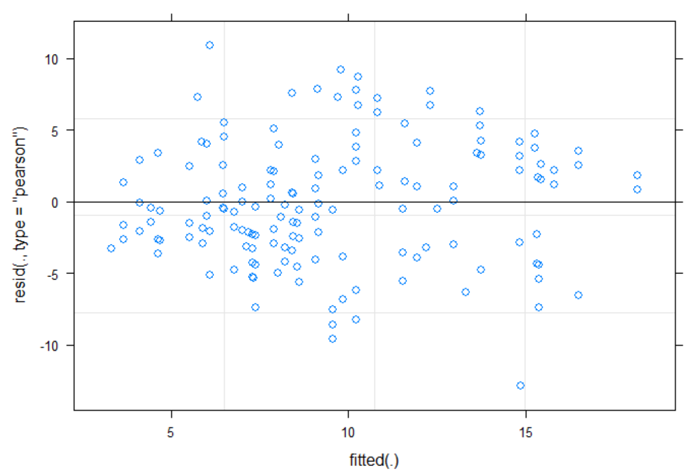

Figure 42. Scatterplot of the fitted Acknowledge Errors model against its residuals.

The model of the effect of the feedback condition, primary role, the frequency of prior video game experience, the frequency of prior video game experience, the level of self-reported feedback use, and the level of teammate familiarity on Assess Errors was evaluated for normality of errors (Figure 43) and the normality of the residuals (Figure 44).

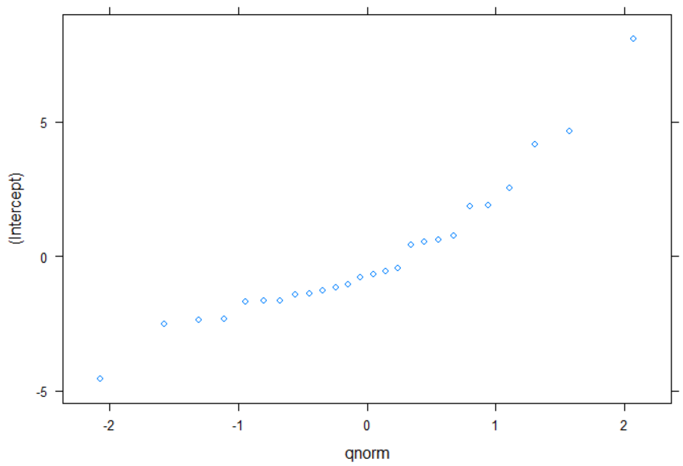

Figure 43. Quantile plot of the random effect of team for the Assess Errors model.

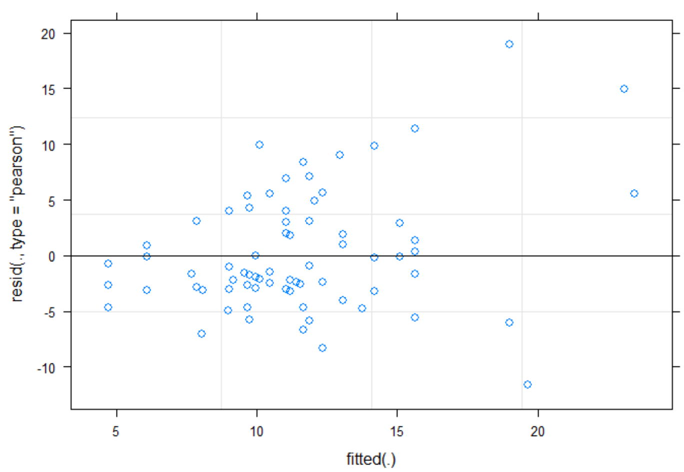

Figure 44. Scatterplot of the fitted Assess Errors model against its residuals.

APPENDIX C. IRB 15-399

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
2420 Lincoln Way, Suite 202
Ames, Iowa 50014
515 294-4566

Date: 03/23/2018

To: Stephen Gilbert

From: Office for Responsible Research

Title: **Team Training in Virtual Environments**

IRB ID: **15-399**

Submission Type: Modification **Review Type**: Full Committee

Approval Date: 03/20/2018 **Date for Continuing Review: 07/20/2018**

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study. Approval is granted with the following condition:

Before the activities described in this application can be initiated, approval must be obtained from the Human Research Protection Official (HRPO) within the Department of Defense.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, including the **recruitment materials and informed consent documents that have the IRB approval stamp**.
- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required.
- **Obtain IRB approval prior to implementing any changes** to the study.
- **Inform the IRB if the Principal Investigator and/or Supervising Investigator end their role or involvement with the project** with sufficient time to allow an alternate PI/Supervising Investigator to assume oversight responsibility. Projects must have an eligible PI to remain open.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- **Stop all human subjects research activity if IRB approval lapses**, unless continuation is necessary to prevent harm to research participants. Human subjects research activity can resume once IRB approval is re-established.

APPENDIX D. SHARED SITUATIONAL AWARENESS QUIZZES

The following quizzes were used to assess participant understanding of the goals of each of the roles (Sniper and Spotter) as well as the predefined task sequence. The Spotter Goal Awareness quiz was only used in the Shared Role Understanding measure, which was created by coding each participants answer not as correct (1) or incorrect (0), but as matching both teammates (2), matching one teammate (1), or matching no teammates (0). All scores were divided by the total possible score to create a percentage correct.

The Team Task Awareness quiz was scored using a Spearman rank correlation of each participant's response to the key of correct answers, which was generated according to the information taught in the tutorial video.

The correct answers to all three of these quizzes, as compared to a key based on the tutorial video, are included alongside the original quizzes themselves, which are featured below. These quizzes were edited after an analysis of internal consistency to produce the versions which are detailed in the text, above.

Sniper Goal Awareness

For the following questions, think back to the two main roles (Spotter and Sniper) that were present in the task. Select as many items that apply to the following questions.

What are the Goals of the Sniper in this Task?

☐ To identify targets new to their zone

☐ To identify targets leaving their zone

☐ To locate targets in their zone

☑ To assess the treats posed by targets

☑ To acknowledge what their teammates say

☐ To keep count of how many targets have left and entered their zone

☐ To keep count of how many OPFOR are on the map

☐ To keep count of how many civilians are on the map

☐ To count the number of OPFOR wearing vests

Spotter Goal Awareness

What are the Goals of the Spotters in this Task?

☑ To identify targets new to their zone

☐ To identify targets leaving their zone

☑ To locate targets in their zone

☐ To assess the treats posed by targets

☑ To acknowledge what their teammates say

☐ To keep count of how many targets have left and entered their zone

☐ To keep count of how many OPFOR are on the map

☐ To keep count of how many civilians are on the map

☐ To count the number of OPFOR wearing vests

Team Task Awareness

For the following question, imagine a scenario in which a single target is crossing between zones.

Please order the following steps in the order that they would be completed when transferring a single target in this task:

Steps	**True to the task**	**Not true to the task**
Spotter 1 sees a target approaching the 1 pole	1	
Spotter 1 transfers a target by pressing the 1 key	2	
Spotter 1 transfers a target by pressing the E key		X
Spotter 2 acknowledges his/her teammate's communication by pressing the E key	3	
Spotter 2 acknowledges his/her teammate's communication by pressing the 1 key		X
Spotter 2 sees a target by the 1 pole	4	
Spotter 2 identifies that a target has entered his/her zone by pressing the SPACEBAR key	5	
Spotter 2 identifies that a target has entered his/her zone by pressing the E key		X
Spotter 2 identifies that a target has entered his/her zone by pressing the 1 key		X
Spotter 2 informs Sniper that a target has entered his/her zone	6	
Sniper acknowledges his/her teammate's communication by pressing the E key	7	
Sniper searches for a target in Spotter 2's zone	8	

Steps	True to the task	Not true to the task
in the direction of the 1 pole		
Sniper spots a target and assesses the threat level posed by the target	9	
Sniper believes target to be a civilian and presses the C key		X
Sniper believes target to be a civilian and presses the X key		X
Sniper believes target to be a civilian and presses the Z key	10	